

A Hierarchy of Information Quantities for Finite Block Length Analysis of Quantum Tasks

Marco Tomamichel and Masahito Hayashi

Abstract—We consider two fundamental tasks in quantum information theory, data compression with quantum side information as well as randomness extraction against quantum side information. We characterize these tasks for general sources using so-called one-shot entropies. These characterizations—in contrast to earlier results—enable us to derive tight second order asymptotics for these tasks in the i.i.d. limit. More generally, our derivation establishes a hierarchy of information quantities that can be used to investigate information theoretic tasks in the quantum domain: The one-shot entropies most accurately describe an operational quantity, yet they tend to be difficult to calculate for large systems. We show that they asymptotically agree (up to logarithmic terms) with entropies related to the quantum and classical information spectrum, which are easier to calculate in the i.i.d. limit. Our technique also naturally yields bounds on operational quantities for finite block lengths.

Index Terms—source compression, randomness extraction, finite block length, second order asymptotics, one-shot entropies, information spectrum

I. INTRODUCTION

THE characterization of information theoretic tasks that are repeated only once (the *one-shot* setting) or a finite number of times (the *finite block length* setting) has recently generated great interest in classical information theory [1], [2]. In particular, these studies investigate the asymptotic performance of information theoretic tasks in the second order, i.e., they determine precisely the contribution to the rate that is proportional to $1/\sqrt{n}$ when we consider n independent and identically distributed (i.i.d.) repetitions of a task. Among the tasks that have been studied in this way are noiseless source coding [3], [4], Slepian-Wolf coding [5], [6], random number generation when the source distribution is known [4], the classical statistical evaluation used for parameter estimation in quantum cryptography [7], and channel coding [1], [2], [8].

Concurrently, progress has been made towards characterizing tasks utilizing quantum resources in the same setting. Two different, but related [9], techniques have been proposed to achieve this: one-shot entropies [10] and a quantum generalization [11], [12] of the information spectrum method [13], [14]. The one-shot approach provides bounds on operational quantities in terms of entropies for general sources. These can be computed for small examples, but are generally difficult to calculate even in the i.i.d. case. We relate these entropies

to the information spectrum of the source, which can be approximated in the i.i.d. setting [4] to yield an asymptotic second order expansion. Combining the two techniques, we thus derive a second order expansion of operational quantities. This is the first such expansion in the quantum regime.

We first give a brief overview of the two techniques.

A. One-Shot Entropies

Motivated by classical cryptography, Renner and coworkers generalized the *min-entropy* (i.e. the Rényi entropy [15] of order ∞) to the quantum setting and used it to investigate randomness extraction against quantum adversaries [16], [10]¹. Together with a technique called *smoothing*, i.e. an optimization of the min-entropy over close states, this result implies a direct bound on randomness extraction against quantum side information, which is tight in the first order [10]. Subsequently, the smooth entropy framework has been refined [17], [18], [19] and used to characterize other tasks in quantum information theory, for example source coding [20], state merging [21], and quantum channel coding [22].

Generally, these results consider *operational quantities* in the one-shot setting and provide direct and converse bounds on them that are valid for general sources and channels. In this work, given a source that emits a random variable, X , and (potentially quantum) side information, B , about X , the following two operational quantities are considered.

- The maximal number of random and secret bits, ε -close to uniform and independent of B , that can be extracted from X is denoted $\ell^\varepsilon(X|B)$. This task was first investigated by Renner and König [16] in the quantum setting and has various applications in cryptography.
- The minimal length in bits of an encoding M of X , such that X can be recovered up to an error ε from B and M is denoted $m^\varepsilon(X|B)$. This is noiseless source compression with side information and has been investigated by Devetak and Winter [23] in the quantum setting.

Direct and converse bounds on these operational quantities are then given in terms of quantities we henceforth call *one-shot entropies*, of which Renner's smooth min-entropy, $H_{\min}^\varepsilon(X|B)$, is the most prominent example. These entropies exhibit useful monotonicity properties, similar to those of the Shannon entropy. Moreover, they can be evaluated numerically for small examples and asymptotically converge to the conditional von Neumann entropy.

¹This is also known as the Leftover Hash Lemma in cryptography.

M. Tomamichel is with the Centre of Quantum Technologies, National University of Singapore, 3 Science Drive 2, Singapore 117542. (email: cqtmarco@nus.edu.sg). M. Hayashi is with the Graduate School of Mathematics, Nagoya University, Furocho, Chikusa-ku, Nagoya, 464-860, Japan, and with the Centre for Quantum Technologies, National University of Singapore, 3 Science Drive 2, Singapore 117542. (email: masahito@math.nagoya-u.ac.jp).

For the operational quantities mentioned above, their one-shot characterization in [10] and [20] directly implies that [19]

$$\begin{aligned} \ell^\varepsilon(X^n|B^n) &= nH(X|B) + O(\sqrt{n}) \quad \text{and} \\ m^\varepsilon(X^n|B^n) &= nH(X|B) + O(\sqrt{n}), \end{aligned} \quad (1)$$

where $0 < \varepsilon < 1$ and the operational quantities are evaluated for n i.i.d. copies of the source. This should be read as follows: independent of the allowed error ε , the tasks can be achieved if and only if the rate is below the Shannon limit, $H(X|B)$. Note that these asymptotic results can be proven more directly, as has been done in [23] for source coding with side information. However, in addition to the asymptotic results, the one-shot approach often naturally yields bounds for finite block lengths, i.e., explicit expressions for the terms $O(\sqrt{n})$ for finite n .

B. The Information Spectrum Method

The information spectrum technique has been introduced by Han and Verdú [13] as a method to treat general information sources beyond the i.i.d. scenario. Han succeeded in treating many major topics in information theory from a unified viewpoint [14] by describing the asymptotic optimal performance in terms of the asymptotic stochastic behavior of the logarithmic likelihood ratio and the logarithmic likelihood. Recently, the information spectrum method was employed to analyze the second order asymptotics of various tasks [4], most prominently channel coding [2].

1) *Classical Information Spectrum*: Given a probability distribution, $P(x)$, and a second, not necessarily normalized, distribution, $Q(x)$, the logarithmic likelihood ratio is the random variable $Z = \log P(X) - \log Q(X)$ under the probability distribution P . To investigate tasks in the i.i.d. limit in first order, it is often sufficient to investigate the expectation value of the likelihood ratio, which evaluates to the relative entropy, $\mathbb{E}[Z] = D(P\|Q) = \sum_x P(x)(\log P(x) - \log Q(x))$.

However, going beyond i.i.d. sources, the information spectrum method is concerned with the full spectrum of this random variable. In this work, we thus introduce the *classical entropic information spectrum*,

$$D_s^\varepsilon(P\|Q) = \sup\{R \in \mathbb{R} \mid P\{Z \leq R\} \leq \varepsilon\}. \quad (2)$$

(The relation of this quantity to a more traditional formulation of the information spectrum is discussed in Section VIII.)

The following crucial observation highlights the usefulness of the classical information spectrum to derive a second order expansion. We consider n -fold i.i.d distributions $P^n(\vec{x}) = \prod_i P(x_i)$ and $Q^n(\vec{x}) = \prod_i Q(x_i)$. Then, the classical entropic information spectrum evaluates to

$$D_s^\varepsilon(P^n\|Q^n) = \sup\left\{R \in \mathbb{R} \mid P^n\left\{\sum Z_i \leq R\right\} \leq \varepsilon\right\},$$

where $Z_i = \log P(X_i) - \log Q(X_i)$ and we are thus left with the spectrum of a sum of i.i.d. random variables. In this case, the law of large numbers implies that the distribution of

$$\frac{\sum_i Z_i - n\mu}{\sqrt{n}\sigma}, \quad \text{where } \mu = \mathbb{E}[Z] \quad \text{and} \quad \sigma = \sqrt{\mathbb{E}[(Z - \mu)^2]},$$

converges to the normal distribution. More precisely, the Berry-Esseen theorem implies that

$$D_s^\varepsilon(P^n\|Q^n) = nD(P\|Q) + \sqrt{n}s(P\|Q)\Phi^{-1}(\varepsilon) + O(1),$$

where $\sigma = s(P\|Q)$ is the standard deviation of the logarithmic likelihood ratio and Φ is the cumulative normal distribution. The information spectrum thus has a natural second order expansion in the i.i.d. limit. (This derivation is presented in more detail in Section VI.)

2) *Quantum Information Spectrum*: A first quantum extension of the information spectrum was investigated by Nagaoka and one of the present authors [11], [12] in order to treat classical-quantum (CQ) channel coding and hypothesis testing for general sequences of channels and sources. In [12], they also clarified the relation between CQ channel coding and quantum hypothesis testing (see also [24], [25], [26], [27] for important contributions to quantum hypothesis testing). The *quantum entropic information spectrum* is denoted $D_s^\varepsilon(\rho\|\sigma)$, and reduces to (2) if ρ and σ commute. (We will define it in Section V.)

However, in contrast to the classical information spectrum, its quantum extension is difficult to calculate or approximate, even in the i.i.d. limit, due to the non-commutativity of ρ and σ . A potential remedy was proposed by Hiai and Petz [26] in the context of hypothesis testing. They considered the joint measurement of σ and $\mathcal{E}_\sigma(\rho)$, where the latter is modified from ρ by pinching in the spectral decomposition of σ . In [28], these two density operators were then used to introduce an alternative quantum extension of the information spectrum, $D_s^\varepsilon(\mathcal{E}_\sigma(\rho)\|\sigma)$. Since the operators σ and $\mathcal{E}_\sigma(\rho)$ commute, this quantum version can be treated similar to the information spectrum of two classical distributions. Moreover, Hiai and Petz [26] showed that $D(\mathcal{E}_{\sigma^n}(\rho^n)\|\sigma^n) = nD(\rho\|\sigma) + o(n)$ for i.i.d. product states $\rho^n = \rho^{\otimes n}$ and $\sigma^n = \sigma^{\otimes n}$. This analysis was generalized [28] to show that $D_s^\varepsilon(\mathcal{E}_{\sigma^n}(\rho^n)\|\sigma^n) = D(\rho\|\sigma) + o(n)$. However, a major drawback of this definition is that $\mathcal{E}_{\sigma^n}(\rho^n)$ is generally not i.i.d. even if ρ^n and σ^n are i.i.d. product states, which makes a second order evaluation of the information spectrum in the asymptotic limit difficult.

Furthermore, in order to show the converse part of the quantum Chernoff bound in hypothesis testing, Nussbaum and Szkoła [29] introduced a pair of distributions related to ρ and σ , which inherit the i.i.d. structure of ρ and σ and have the convenient property that the first two moments of their likelihood ratio coincides with the first two moments of the likelihood ratio of ρ and σ . Using the eigenvalue decompositions $\rho = \sum_x r_x |v_x\rangle\langle v_x|$ and $\sigma = \sum_y s_y |u_y\rangle\langle u_y|$, these distributions are given by

$$P_{\rho,\sigma}(x,y) = r_x |\langle v_x | u_y \rangle|^2 \quad \text{and} \quad Q_{\rho,\sigma}(x,y) = s_y |\langle v_x | u_y \rangle|^2,$$

and their entropic information spectrum, $D_s^\varepsilon(P_{\rho,\sigma}, Q_{\rho,\sigma})$, will play an important role in our analysis.

C. Main Results

In this work, we improve the analysis leading to the asymptotic expansion of the operational quantities in Eq. (1), relying on techniques developed for the smooth entropy framework as well as the quantum information spectrum method.

1) Second Order Expansion of Operational Quantities:

Our first contribution is to show that both the direct and converse bounds on the operational quantities converge to the same expression in the second order. In particular, we find the following asymptotic expansion for n i.i.d. copies of the source and $0 < \varepsilon < 1$:

$$\begin{aligned}\ell^\varepsilon(X^n|B^n) &= nH(X|B) + \sqrt{n}s(X|B)\Phi^{-1}(\varepsilon^2) + O(\log n), \\ m^\varepsilon(X^n|B^n) &= nH(X|B) - \sqrt{n}s(X|B)\Phi^{-1}(\varepsilon) + O(\log n),\end{aligned}$$

where $H(X|B)$ is the conditional von Neumann entropy of the source, Φ is the cumulative normal distribution function, and s describes the standard deviation corresponding to the conditional von Neumann entropy. (See Corollaries 16 and 15.) Note, in particular, that $\Phi^{-1}(\varepsilon)$ is negative for small ε and changes sign when ε exceeds $1/2$. To the best of our knowledge, this constitutes the first second order expansion of an operational quantity involving quantum resources.

2) *Finite Block Length Analysis:* Our analysis naturally yields both direct and converse bounds on the operational quantities for finite n , which can be evaluated numerically.

We give an example of such a finite block length analysis in Figure 1. For this purpose, we consider the state that results when transmitting either $|0\rangle$ or $|1\rangle$ through the complementary channel of a Pauli channel with a phase error $p = 0.05$ that is independent of the bit flip error. The resulting state is

$$\begin{aligned}\rho_{XB} &= \frac{1}{2}|0\rangle\langle 0| \otimes |\phi^0\rangle\langle \phi^0| + \frac{1}{2}|1\rangle\langle 1| \otimes |\phi^1\rangle\langle \phi^1|, \quad \text{where} \\ |\phi^x\rangle &= \sqrt{p}|0\rangle + (-1)^x\sqrt{1-p}|1\rangle.\end{aligned}$$

We are interested in the rate $r = \frac{1}{n}\ell^\varepsilon(X^n|B^n)$ at which uniform and independent randomness can be extracted from X if we require that $\varepsilon = 10^{-6}$. In the first order, the rate approaches $H(X|B) = 1 - h(p) \approx 0.714$. The deviation from this bound in the second order is significant for small n . We have $s(X|B)\Phi^{-1}(\varepsilon^2) \approx -9.6$ for this example, which leads to a drop of 10% in the rate at $n \approx 1.8 \cdot 10^4$. Converse bounds for finite n are relevant as they allow us to investigate how close the performance of a given protocol is to the maximal achievable rate. From Fig. 1, we can deduce that it is impossible to securely extract more than 95% of the Shannon entropy for $n = 10^4$. The calculations leading to the direct and converse bounds on $\ell^\varepsilon(X^n|B^n)_\rho$ for this example are discussed in Appendix A.

3) *Tight One-Shot Characterization:* In order to prove our results, we first find suitable characterizations for ℓ^ε and m^ε in terms of one-shot entropies. We bound ℓ^ε in terms of the smooth min-entropy, $H_{\min}^\varepsilon(X|B)$ of the source, and use a conditional version of the hypothesis testing entropy [30], $H_h^\varepsilon(X|B)$, to bound m^ε . This results in the following bounds, for any $0 < \eta \leq \varepsilon < 1$ (cf. Theorem 8 and 9):

$$\begin{aligned}H_{\min}^{\varepsilon-\eta}(X|B) - \log \frac{1}{\eta^4} - 3 &\leq \ell^\varepsilon(X|B) \leq H_{\min}^\varepsilon(X|B), \\ H_h^\varepsilon(X|B) &\leq m^\varepsilon(X|B) \leq H_h^{\varepsilon-\eta}(X|B) + \log \frac{\varepsilon}{\eta^2} + 3.\end{aligned}\quad (3)$$

Here, ε is a parameter of the problem and is kept constant, whereas η can be optimized over.

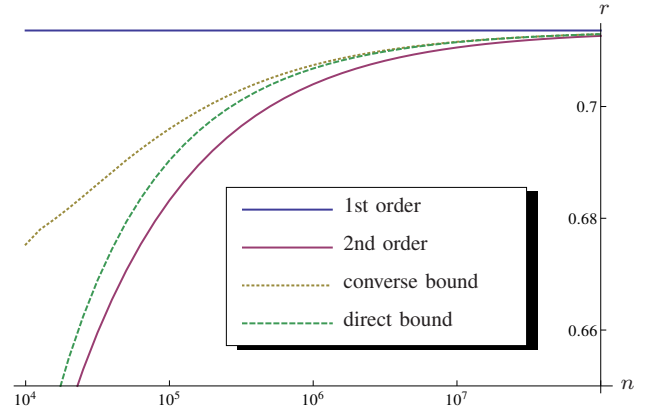


Fig. 1. This plot shows direct and converse bounds on the extraction rate for $n \in [10^4, 10^8]$. For increasing n the bounds first converge to each other, then to the second order asymptotics, and finally to the Shannon rate.

These one-shot characterizations are tighter than earlier results in [10], [20]. In particular, even if we choose $\eta = \text{poly}(n)^{-1}$, the additive terms grow at most as $\log n$ for large enough n . The smoothing parameter of the one-shot entropies (in both bounds) thus inherits the operational interpretation of ε as the allowed error. This is of crucial importance as the second order expansion we aim to show is a function of ε , and we thus need to find lower and upper bounds in terms of the same parameter.

4) *Hierarchy of Information Quantities:* Furthermore, we establish a hierarchy of information quantities (cf. Figure 2) that can be used to analyze quantum information tasks beyond the examples discussed above. The hierarchy is partly inspired by recent results in hypothesis testing and constitutes the main technical contribution of this paper.

The operational quantities (Class 1) describe the optimal achievable performance of a task, and they depend strongly on the exact specification of the considered task. For example, in the case of randomness extraction, this quantity depends on the precise security requirement we impose on the extracted random variable. To calculate these quantities, one needs to optimize the performance over all valid protocols, which is difficult even for small examples. In a first step, we thus bound the operational quantities in terms of one-shot entropies (Class 2), in our case H_{\min}^ε and H_h^ε . These quantities can be formulated as semi-definite optimization problems (SDPs)² and can therefore be calculated for small examples.

In the i.i.d. setting for large block lengths (e.g., $n \gg 10$), however, these optimization problems quickly become intractable as their complexity scales exponentially in n . Thus, we relate the one-shot entropies to the quantum information spectrum (Class 3) and then the classical information spectrum of the corresponding Nussbaum-Szkoła distributions (Class 4), which can often be approximated even for large n . This can be done incurring an additive error term that scales at most logarithmically in n . Finally, the classical entropic information

²The min-entropy can be formulated as an SDP [31] and an extension of this to the smooth min-entropy is possible. The SDP for hypothesis testing is discussed in Section II.

Class	Role	Quantities
Class 1	Describes the optimal achievable performance. The calculation is very difficult, even for small examples.	$m^\varepsilon(X B)$ $l^\varepsilon(X B)$,
Class 2	Bound for general sources in terms of one-shot entropies. The calculation is possible for small examples, using an SDP.	$H_h^{\varepsilon \pm \eta}(X B)$, $H_{\min}^{\varepsilon \pm \eta}(X B)$, $D_h^{\varepsilon \pm \eta}(\rho \sigma)$, $D_{\max}^{\varepsilon \pm \eta}(\rho \sigma)$
Class 3	Quantum version of the entropic information spectrum.	$D_s^{\varepsilon \pm \delta}(\rho \sigma)$, $D_s^{\varepsilon \pm \delta}(\mathcal{E}_\sigma(\rho) \sigma)$
Class 4	Classical entropic information spectrum. The calculation is approximately possible for i.i.d. sources.	$D_s^{\varepsilon \pm \delta}(P_{\rho,\sigma} Q_{\rho,\sigma})$
Class 5	Asymptotic expansion in the second order. The calculation is easy for arbitrarily large n .	$nD(\rho \sigma) + \sqrt{n}s(\rho \sigma)\Phi^{-1}(\varepsilon)$ $nH(X B) + \sqrt{n}s(X B)\Phi^{-1}(\varepsilon)$

Classes	Difference	Method
Class 1 \rightarrow Class 2	$O(\log n)$ with $\eta \propto \frac{1}{\sqrt{n}}$	One-shot analysis, random coding and monotonicity (cf. Thms. 9 and 8).
Class 2 \rightarrow Class 4	$O(\log n)$ with $\delta \propto \frac{1}{\sqrt{n}}$	Relations between relative entropies (cf. Prop. 13 and Thm. 14).
Class 4 \rightarrow Class 5	$O(1)$	Berry-Esseen Theorem (cf. Eq. (30)).

Fig. 2. Hierarchy of information quantities. We consider operational quantities for a constant $\varepsilon \in (0, 1)$ and n i.i.d. uses of the source. The one-shot entropies (Class 2) provide a “microscopic” analysis of the optimal performance of a task (Class 1) for general sources, whereas the information spectrum (Classes 3 and 4) and their asymptotic expansion (Class 5) give a “macroscopic” view that can be approximately calculated for sources with sufficient structure. (Note that Class 3 and Class 4 are unified if ρ and σ commute.)

spectrum allows us to evaluate the second order asymptotic expansion precisely (Class 5).

Formally, we show that the following quantities are equivalent in an appropriate sense (cf. Proposition 13 and Theorem 14):

$$D_{\max}^{\sqrt{1-\varepsilon}}(\rho||\sigma) \approx D_h^\varepsilon(\rho||\sigma) \approx D_s^\varepsilon(\rho||\sigma) \\ \approx D_s^\varepsilon(\mathcal{E}_\sigma(\rho)||\sigma) \approx D_s^\varepsilon(P_{\rho,\sigma}||Q_{\rho,\sigma}).$$

First, note that ε is varied by an additive optimization parameter in some relations, analogously to the situation in Eq. (3). More importantly, the equivalence only holds up to additive terms $\log \theta(\sigma)$, where $\theta(\sigma)$ is the logarithm of the ratio between the largest and smallest eigenvalue of σ . In the i.i.d. setting, it is evident that $\theta(\sigma^n)$ grows at most linearly in n and this additive term thus grows at most as $O(\log n)$. Hence, our results imply that the smoothing parameter for all these quantities can be chosen as $\varepsilon \pm \text{poly}(n)^{-1}$ without incurring a penalty that grows faster than $O(\log n)$.

5) *Convergence to Relative Entropy*: Our analysis also improves on earlier work [10], [17], [32], [19] that investigated the convergence of one-shot entropies in the i.i.d. setting for finite n . Given a quantum state ρ and an a positive semi-definite operator σ , these earlier results imply that the i.i.d. product states ρ^n and σ^n satisfy

$$D_{\max}^\varepsilon(\rho^n||\sigma^n) = nD(\rho||\sigma) + O(\sqrt{n}), \quad \text{and} \\ D_h^\varepsilon(\rho^n||\sigma^n) = nD(\rho||\sigma) + O(\sqrt{n}),$$

where D_h^ε is the hypothesis testing entropy and D_{\max}^ε a relative entropy version of the smooth min-entropy. These results also establish explicit upper and lower bounds on the convergence for finite n , however, the second order term, scaling as \sqrt{n} , is not tight. Our analysis implies improved bounds for finite n , which are tight in the second order. We find

$$D_{\max}^\varepsilon(\rho^n||\sigma^n) = nD(\rho||\sigma) - \sqrt{n}s(\rho||\sigma)\Phi^{-1}(\varepsilon^2) + O(\log n), \\ D_h^\varepsilon(\rho^n||\sigma^n) = nD(\rho||\sigma) + \sqrt{n}s(\rho||\sigma)\Phi^{-1}(\varepsilon) + O(\log n).$$

(These statements are shown in Section VI-A.)

We also point the reader to independent and concurrent work by Li [33], who also recovers the i.i.d. second order asymptotics for hypothesis testing using a more direct approach.

D. Outline

The remainder of the paper is structured as follows. In Section II, we introduce the notation and the one-shot entropies required for our discussion. We show some properties of these one-shot entropies, which we will need for subsequent proofs. In Sections III and IV we characterize source compression and randomness extraction with quantum side information in the one-shot setting, respectively. Section V then introduces relations between different information quantities which are used extensively to derive our asymptotic results. Section VI is devoted to the analysis of the two operational tasks in the i.i.d. asymptotic setting and Section VII covers finite block lengths. Finally, Section VIII illuminates the relation between the traditional formulation of the information spectrum method and its entropic version used in this work.

II. PRELIMINARIES

A. Notation and Definitions

For a finite-dimensional Hilbert space \mathcal{H} , we denote by $\mathcal{L}(\mathcal{H})$ and $\mathcal{P}(\mathcal{H})$ linear and positive semi-definite operators on \mathcal{H} , respectively. Quantum states are in the set $\mathcal{S}(\mathcal{H}) = \{\rho \in \mathcal{P}(\mathcal{H}) : \text{tr}(\rho) = 1\}$ and we also define the set of sub-normalized quantum states, $\mathcal{S}_{\leq}(\mathcal{H}) = \{\rho \in \mathcal{P}(\mathcal{H}) : 0 < \text{tr}(\rho) \leq 1\}$. On $\mathcal{L}(\mathcal{H})$, we employ the Schatten ∞ -norm $\|\cdot\|$, which evaluates to the largest singular value, and the Hilbert-Schmidt inner product $\langle L, M \rangle = \text{tr}(L^\dagger M)$.

We write $A \geq B$ if and only if $A - B \in \mathcal{P}(\mathcal{H})$. When comparing a scalar to a matrix, we implicitly assume that it is multiplied by the identity matrix, which we denote by 1. We use $\{A \geq B\}$ to denote the projector onto the space spanned by the eigenvectors of $A - B$ that corresponds to non-negative

eigenvalues. Clearly, this definition implies that $\{A \geq B\}A \geq \{A \geq B\}B$. Moreover, we have $\{A > B\} = 1 - \{A \leq B\}$.

Multipartite quantum systems are described by tensor product spaces. We use capital letters to denote the different systems and subscripts to indicate on what subspace an operator acts. For example, if L_{AB} is an operator on $\mathcal{H}_{AB} = \mathcal{H}_A \otimes \mathcal{H}_B$, then $L_A = \text{tr}_B(L_{AB})$ is implicitly defined as its marginal on A . Conversely, we call L_{AB} an extension of L_A . We call a state classical-quantum (CQ) if it is of the form $\rho_{XA} = \sum_x p_x |x\rangle\langle x| \otimes \phi_A^x$, where $\phi_A^x \in \mathcal{S}(\mathcal{H}_A)$, p_x a probability distribution, and $\{|x\rangle\}$ an orthonormal basis of \mathcal{H}_X . We call X a register to distinguish it from genuinely quantum systems.

A map $\mathcal{E} : \mathcal{L}(\mathcal{H}) \rightarrow \mathcal{L}(\mathcal{H}')$ is called a completely positive map (CPM) if it maps $\mathcal{P}(\mathcal{H} \otimes \mathcal{H}'')$ into $\mathcal{P}(\mathcal{H}' \otimes \mathcal{H}'')$ for any \mathcal{H}'' . It is called trace-preserving (TP) if $\text{tr}(\mathcal{E}(X)) = \text{tr}(X)$ for any $X \in \mathcal{P}(\mathcal{H})$. It is called unital if $\mathcal{E}(1) = 1$ and sub-unital if $\mathcal{E}(1) \leq 1$. The adjoint map \mathcal{E}^\dagger of a map \mathcal{E} is defined via the relation $\langle \mathcal{E}(L), M \rangle = \langle L, \mathcal{E}^\dagger(M) \rangle$ for all L, M . Adjoint maps of CPMs are CPMs. Moreover, adjoint maps of TP-CPMs are unital CPMs and vice versa. As an example, consider the pinching TP-CPM $\mathcal{M}_X : L \mapsto \sum_x |x\rangle\langle x| L |x\rangle\langle x|$ and note that this map is self-adjoint with regards to the Hilbert Schmidt inner product and, thus, also unital.

The following result generalizes pinching.

Lemma 1: Let $A \in \mathcal{P}(\mathcal{H})$ and let $\mathcal{E} : \mathcal{L}(\mathcal{H}) \rightarrow \mathcal{L}(\mathcal{H}')$ be a sub-unital CPM. Then, $\|\mathcal{E}(A)\| \leq \|A\|$.

Proof: We may write the Schatten infinity norm (of positive semi-definitive operators) as an SDP

$$\|A\| = \max_{Q \geq 0} \text{tr}(AQ) = \min_{A \leq \gamma 1} \gamma.$$

Thus, if γ^* is minimal for $\|A\|$, we have $\mathcal{E}(A) \leq \gamma^* \mathcal{E}(1) \leq \gamma^*$. Hence, γ^* is feasible for $\|\mathcal{E}(A)\|$, concluding the proof. ■

B. The Smooth Entropy Framework

We use the purified distance [18] on sub-normalized quantum states and write $\rho \approx^\varepsilon \sigma$ if and only if $P(\rho, \sigma) \leq \varepsilon$, where $P(\rho, \sigma) := \sqrt{1 - F^2(\rho, \sigma)}$ and

$$F(\rho, \sigma) := \text{tr}|\sqrt{\rho}\sqrt{\sigma}| + \sqrt{(1 - \text{tr}\rho)(1 - \text{tr}\sigma)}.$$

generalizes the fidelity to sub-normalized states. The purified distance has the following properties [18].

- Uhlmann's theorem: Let $\rho_{AB} \in \mathcal{S}_{\leq}(\mathcal{H}_{AB})$ be a bipartite state and $\tau_A \in \mathcal{S}_{\leq}(\mathcal{H}_A)$ with $\tau_A \approx^\varepsilon \rho_A$. Then, there exists an extension τ_{AB} of τ_A such that $\tau_{AB} \approx^\varepsilon \rho_{AB}$.
- Monotonicity: Let \mathcal{E} be a trace non-increasing CPM. Then, $\rho \approx^\varepsilon \tau \implies \mathcal{E}(\rho) \approx^\varepsilon \mathcal{E}(\tau)$.
- Fuchs-van de Graaf: $D(\rho, \tau) \leq P(\rho, \tau) \leq \sqrt{2D(\rho, \tau)}$, where $D(\rho, \tau) = \frac{1}{2} \text{tr}|\rho - \tau| + \frac{1}{2} |\text{tr}\rho - \text{tr}\tau| \leq \text{tr}|\rho - \tau|$.

For any sub-normalized state $\rho \in \mathcal{S}_{\leq}(\mathcal{H})$, we use $\mathcal{B}^\varepsilon(\rho) := \{\tilde{\rho} \in \mathcal{S}_{\leq}(\mathcal{H}) \mid \tilde{\rho} \approx^\varepsilon \rho\}$ to denote the set of states close to ρ .

We use the following relative and conditional entropies [34], [18], which are based on Renner's initial definition of the smooth min-entropy [10].

Definition 1: Let $\rho \in \mathcal{S}(\mathcal{H})$, $\sigma \in \mathcal{P}(\mathcal{H})$, and $0 \leq \varepsilon < 1$. Then, the relative max-entropy is defined as

$$D_{\max}^\varepsilon(\rho \parallel \sigma) := \min_{\tilde{\rho} \approx^\varepsilon \rho} \inf \{\lambda \in \mathbb{R} \mid \tilde{\rho} \leq 2^\lambda \sigma\}.$$

where we optimize over all $\tilde{\rho} \in \mathcal{S}_{\leq}(\mathcal{H})$ with $\tilde{\rho} \approx^\varepsilon \rho$.

Definition 2: Let $\rho_{AB} \in \mathcal{S}(\mathcal{H}_{AB})$ and $0 \leq \varepsilon < 1$. Then, the smooth min-entropy is defined as

$$H_{\min}^\varepsilon(A|B)_\rho := \max_{\sigma_B} -D_{\max}^\varepsilon(\rho_{AB}, 1_A \otimes \sigma_B),$$

where the optimization is over all $\sigma_B \in \mathcal{S}(\mathcal{H}_B)$.

We employ the following property of this entropy [18].

Lemma 2: Let $\rho_{AB} \in \mathcal{S}_{\leq}(\mathcal{H}_{AB})$ and let $U : A \rightarrow A'$ be an isometry. Then, $H_{\min}^\varepsilon(A|B)_\rho = H_{\min}^\varepsilon(A'|B)_{U\rho U^\dagger}$.

Moreover, the smooth min-entropy is monotonous under the application of classical functions.³

Proposition 3: Let $\rho_{XAB} = \sum_x p_x |x\rangle\langle x| \otimes \phi_{AB}^x$ be a state in $\mathcal{S}_{\leq}(\mathcal{H}_{XAB})$ and let $f : X \rightarrow Z$ be a function. Then,

$$H_{\min}^\varepsilon(XA|B)_\rho \geq H_{\min}^\varepsilon(f(X)A|B)_\rho, \quad \text{where} \\ \rho_{ZAB} = \sum_z |z\rangle\langle z| \otimes \left(\sum_{x \in f^{-1}(z)} p_x \phi_{AB}^x \right).$$

Proof: We first show the statement for the trivial function $f(x) = 1$, i.e. we show that $H_{\min}^\varepsilon(XA|B)_\rho \geq H_{\min}^\varepsilon(A|B)_\rho$.

For a state $\tilde{\rho}_{AB} \approx^\varepsilon \rho_{AB}$ that maximizes $H_{\min}^\varepsilon(A|B)_\rho$, we define an extension $\tilde{\rho}_{XAB} \in \mathcal{S}_{\leq}(\mathcal{H}_{XAB})$ with $\tilde{\rho}_{XAB} \approx^\varepsilon \rho_{XAB}$ using Uhlmann's theorem for the purified distance. Furthermore, using the pinching $\mathcal{M}_X : L \mapsto \sum_x |x\rangle\langle x| L |x\rangle\langle x|$, we define

$$\hat{\rho}_{XAB} := \mathcal{M}_X(\tilde{\rho}_{XAB}) \approx^\varepsilon \mathcal{M}_X(\rho_{XAB}) = \rho_{XAB}.$$

Now, note that $\hat{\rho}_{XAB}$ has the form $\hat{\rho}_{XAB} = \sum_x \hat{p}_x |x\rangle\langle x| \otimes \hat{\phi}_{AB}^x$ and $\hat{\rho}_{AB} = \sum_x \hat{p}_x \hat{\phi}_{AB}^x \leq 2^{-\lambda} 1_A \otimes \sigma_B$ for $\lambda = H_{\min}^\varepsilon(A|B)_\rho$ and some $\sigma_B \in \mathcal{S}(\mathcal{H}_B)$ by definition. Moreover,

$$\forall x : \hat{p}_x \hat{\phi}_{AB}^x \leq 2^{-\lambda} 1_A \otimes \sigma_B \implies \hat{\rho}_{XAB} \leq 2^{-\lambda} 1_{XA} \otimes \sigma_B$$

and we thus have $H_{\min}^\varepsilon(XA|B)_\rho \geq \lambda$, concluding the proof for this special case.

For general functions f , we first apply the isometry $U_f : |x\rangle \mapsto |x\rangle \otimes |f(x)\rangle$ from X to XZ . We then define the state

$$\tau_{XZAB} = U_f \rho_{XAB} U_f^\dagger = \sum_{x,z} p_x \delta_{z,f(x)} |x\rangle\langle x| \otimes |z\rangle\langle z| \otimes \phi_{AB}^x,$$

which is an extension of both ρ_{XAB} and ρ_{ZAB} . The statement now follows from the relation

$$H_{\min}^\varepsilon(XA|B)_\rho = H_{\min}^\varepsilon(XZA|B)_\tau \geq H_{\min}^\varepsilon(ZA|B)_\tau,$$

where the equality is due to Lemma 2 and the inequality is covered by the special case already shown. ■

C. The Hypothesis Testing Entropy

Another one-shot entropy has been used in hypothesis testing and channel coding [30].

Definition 3: Let $\rho \in \mathcal{S}(\mathcal{H})$, $\sigma \in \mathcal{P}(\mathcal{H})$ and $0 \leq \varepsilon \leq 1$. Then, the ε -hypothesis testing relative entropy is defined as

$$2^{-D_h^\varepsilon(\rho \parallel \sigma)} := \inf \{ \langle Q, \sigma \rangle \mid 0 \leq Q \leq 1 \wedge \langle Q, \rho \rangle \geq 1 - \varepsilon \}.$$

It is easy to verify that the infimum is always attained and $D_h^\varepsilon(\rho \parallel \sigma)$ takes values in $[0, \infty]$ when we supplement the real axis by ∞ and define $-\log 0 \equiv \infty$ for the binary entropy log.

³See also [19]. Renner [10] showed this property for $\varepsilon = 0$.

We will use conditional versions of this entropy, defined as follows.

Definition 4: Let $\rho_{AB} \in \mathcal{S}(\mathcal{H}_{AB})$, $\sigma_B \in \mathcal{S}(\mathcal{H}_B)$ and $0 \leq \varepsilon \leq 1$. The conditional ε -hypothesis testing entropies are defined as $H_h^\varepsilon(A|B)_{\rho|\sigma} := -D_h^\varepsilon(\rho_{AB} \| 1_A \otimes \sigma_B)$ and $H_h^\varepsilon(A|B)_\rho := \max_\sigma H_h^\varepsilon(A|B)_{\rho|\sigma}$.

The following two expressions for $2^{H_h^\varepsilon(A|B)_\rho}$ are equivalent due to the strong duality of semi-definite programs [35].

primal problem	dual problem
min: $\ Q_B\ $	max: $\eta(1 - \varepsilon - \text{tr}(N_{AB}))$
s.t.: $0 \leq Q_{AB} \leq 1$	s.t.: $\eta \geq 0, N_{AB} \geq 0, \sigma_B \geq 0$
$\langle Q_{AB}, \rho_{AB} \rangle \geq 1 - \varepsilon$	$\rho_{AB} \leq \frac{1}{\eta} 1_A \otimes \sigma_B + N_{AB}$
	$\text{tr}(\sigma_B) \leq 1$

Accordingly, we call an operator Q_{AB} that satisfies $0 \leq Q_{AB} \leq 1$ and $\langle Q_{AB}, \rho_{AB} \rangle \geq 1 - \varepsilon$ primal feasible. If it also attains the minimum, we call it primal optimal. Similarly, a triple $\{N_{AB}, \sigma_B, \eta\}$ of positive semi-definite operators is called dual feasible if it satisfies $\rho_{AB} \leq \frac{1}{\eta} 1_A \otimes \sigma_B + N_{AB}$ and $\text{tr}(\sigma_B) \leq 1$. It is called dual optimal if it also attains the maximum.

We now explore some properties of the hypothesis testing entropy. It satisfies a data-processing inequality.

Proposition 4: Let $\rho_{AB} \in \mathcal{S}(\mathcal{H}_{AB})$, let $\mathcal{E} : A \rightarrow A'$ be a sub-unital TP-CPM, and let $\mathcal{F} : B \rightarrow B'$ be a TP-CPM. Then, $\tau_{A'B'} = \mathcal{E} \circ \mathcal{F}(\rho_{AB})$ satisfies $H_h^\varepsilon(A|B)_\rho \leq H_h^\varepsilon(A'|B')_\tau$.

Proof: Let $\{N_{AB}, \sigma_B, \eta\}$ be dual optimal for $H_h^\varepsilon(A|B)_\rho$. Then, applying $\mathcal{E} \circ \mathcal{F}$ on both sides of the inequality $\rho_{AB} \leq \frac{1}{\eta} 1_A \otimes \sigma_B + N_{AB}$ leads to

$$\tau_{AB} \leq \frac{1}{\eta} \mathcal{E}(1_A) \otimes \mathcal{F}(\sigma_B) + \mathcal{E} \circ \mathcal{F}(N_{AB}) \leq \frac{1}{\eta} 1_{A'} \otimes \tilde{\sigma}_{B'} + \tilde{N}_{A'B'},$$

where $\tilde{N}_{A'B'} = \mathcal{E} \circ \mathcal{F}(N_{AB})$ and $\tilde{\sigma}_{B'} = \mathcal{F}(\sigma_B)$. Hence, the triple $\{\tilde{N}_{A'B'}, \tilde{\sigma}_{B'}, \eta\}$ is dual feasible and $H_h^\varepsilon(A'|B')_\tau \geq \eta(1 - \varepsilon - \text{tr}(\tilde{N}_{A'B'})) = H_h^\varepsilon(A|B)_\rho$. ■

The following result implies that structure of the state can be translated into structure of the optimal primal and dual.

Lemma 5: Let $\rho_{AB} \in \mathcal{S}(\mathcal{H}_{AB})$ and let $\mathcal{E} : A \rightarrow A$ be a unital TP-CPM and $\mathcal{F} : B \rightarrow B$ be a TP-CPM such that $\mathcal{E} \circ \mathcal{F}(\rho_{AB}) = \rho_{AB}$. Then, the following holds for the SDP for $H_h^\varepsilon(A|B)_\rho$: (1) If Q'_{AB} is primal optimal then $Q_{AB} = \mathcal{E}^\dagger \circ \mathcal{F}^\dagger(Q'_{AB})$ is primal optimal too. (2) If $\{N'_{AB}, \sigma'_B, \eta\}$ is dual optimal, then $\{N_{AB}, \sigma_B, \eta\}$ with $N_{AB} = \mathcal{E} \circ \mathcal{F}(N'_{AB})$ and $\sigma_B = \mathcal{F}(\sigma'_B)$ is dual optimal too.

Proof: Let Q'_{AB} be primal optimal. The operator Q_{AB} as defined above is feasible since the inner product satisfies $\langle \mathcal{E}^\dagger \circ \mathcal{F}^\dagger(Q'_{AB}), \rho_{AB} \rangle = \langle Q'_{AB}, \mathcal{E} \circ \mathcal{F}(\rho_{AB}) \rangle$ and $Q_{AB} \leq 1_{AB}$ since the maps \mathcal{E}^\dagger and \mathcal{F}^\dagger are unital. Moreover, $\|\mathcal{F}^\dagger(Q_B)\| \leq \|Q_B\|$ due to Lemma 1, which implies that Q_{AB} must be optimal. Similarly, let $\{N'_{AB}, \sigma'_B, \eta\}$ be dual optimal, then $\{N_{AB}, \sigma_B, \eta\}$ as defined above is clearly dual feasible and optimal. ■

Corollary 6: Let $\rho_{XAYB} \in \mathcal{S}(\mathcal{H}_{XAYB})$ be classical on X and Y . Then, the SDP for $H_h^\varepsilon(AX|BY)_\rho$ has a primal optimal operator of the form $Q_{XAYB} = \sum_{x,y} |x\rangle\langle x| \otimes |y\rangle\langle y| \otimes$

Q_{AB}^{xy} and dual optimal operators of the form $N_{XAYB} = \sum_{x,y} |x\rangle\langle x| \otimes |y\rangle\langle y| \otimes N_{AB}^{xy}$ and $\sigma_{YB} = \sum_y |y\rangle\langle y| \otimes \sigma_B^y$.

The following proposition gives bounds on the entropy of classical information.

Proposition 7: Let $\rho_{XAB} = \sum_x p_x |x\rangle\langle x| \otimes \phi_{AB}^x$ be a state in $\mathcal{S}(\mathcal{H}_{XAB})$. Then, using $|X| = |\text{supp}\{p_x\}|$,

$$H_h^\varepsilon(A|B)_\rho \geq H_h^\varepsilon(A|XB)_\rho \geq H_h^\varepsilon(XA|B)_\rho - \log |X|, \quad (4)$$

$$H_h^\varepsilon(A|XB)_\rho \leq H_h^\varepsilon(XA|B)_\rho \leq H_h^\varepsilon(A|B)_\rho + \log |X|. \quad (5)$$

Proof: We prove the four inequalities separately and apply Corollary 6, which ensures that the primal optimal operators are classical on X .

The first inequality in (4) follows from Proposition 4 using the partial trace over X as a TP-CPM.

To prove the second inequality in (4), let $Q_{XAB} = \sum_{x: p_x > 0} |x\rangle\langle x| \otimes Q_{AB}^x$ be primal optimal for $H_h^\varepsilon(A|XB)_\rho = \log \max_x \|Q_{AB}^x\|$. Thus, Q_{XAB} is primal feasible for $H_h^\varepsilon(XA|B)_\rho$ and we find

$$\begin{aligned} H_h^\varepsilon(XA|B)_\rho &\leq \log \|\text{tr}_{XA}(Q_{XAB})\| = \log \left\| \sum_{x: p_x > 0} Q_{AB}^x \right\| \\ &\leq \log \left(|\text{supp}\{p_x\}| \max_x \|Q_{AB}^x\| \right), \end{aligned}$$

which implies the result.

The first inequality of (5) follows in a similar fashion, only this time we choose Q_{XAB} to be primal optimal for $H_h^\varepsilon(XA|B)_\rho$. Clearly,

$$\begin{aligned} H_h^\varepsilon(A|XB)_\rho &\leq \log \max_x \|Q_{AB}^x\| \leq \log \left\| \sum_{x: p_x > 0} Q_{AB}^x \right\| \\ &= H_h^\varepsilon(XA|B)_\rho. \end{aligned}$$

Finally, to prove the second inequality in (5), we consider the SDP for $H_h^\varepsilon(A|B)_\rho = \log \|Q_B\|$ with Q_{AB} primal optimal. Then, the operator $Q_{XAB} = \sum_{x: p_x > 0} |x\rangle\langle x| \otimes Q_{AB}$ is primal feasible for $H_h^\varepsilon(XA|B)_\rho$. In particular, it satisfies

$$\begin{aligned} \text{tr}(Q_{XAB} \rho_{XAB}) &= \sum_{x: p_x > 0} p_x \text{tr}(Q_{AB} \phi_{AB}^x) \\ &= \text{tr}(Q_{AB} \rho_{AB}) \geq 1 - \varepsilon, \end{aligned}$$

where the last inequality follows from the primal feasibility of Q_{AB} . Hence, we have $H_h^\varepsilon(XA|B)_\rho \leq \log \|\text{tr}_{XA}(Q_{XAB})\| = \log |\text{supp}\{p_x\}| + \log \|Q_B\|$, which concludes the proof. ■

III. ONE-SHOT CHARACTERIZATION OF RANDOMNESS EXTRACTION

Given a state $\rho_{XB} = \sum_x p_x |x\rangle\langle x| \otimes \phi_B^x$ that is classical on X , we want to extract a random string, Z , that is independent of the quantum side information B . We consider randomized protocols $\mathcal{P} = \{\mathcal{S}, \mathcal{Z}, p_s, h_s\}$ that consists of a seed, \mathcal{S} , an output set \mathcal{Z} , and for all $s \in \mathcal{S}$, a probability p_s and a hash function $h_s : X \rightarrow \mathcal{Z}$.

The protocol now acts on the initial state $\rho_{XBS} = \rho_{XB} \otimes \sum_s p_s |s\rangle\langle s|$ by applying the function h_s depending on the value s in the register S . This results in $\tau_{ZBS} = \sum_s p_s |s\rangle\langle s| \otimes \tau_{ZB}^s$, where $\tau_{ZB}^s = \sum_z |z\rangle\langle z| \otimes \sum_{x \in h_s^{-1}(z)} p_x \phi_B^x$. Alternatively, this evaluation can be modeled using a TP-CPM \mathcal{F} from XS to ZS , such that $\tau_{ZBS} = \mathcal{F}(\rho_{XBS})$.

We characterize the extractable randomness of a state ρ_{XB} by the amount of randomness (in bits) that can be extracted such that Z is close to uniform and independent of E and S .

Definition 5: Let $\rho_{XB} \in \mathcal{S}(\mathcal{H}_{XB})$ be a CQ state. Then, we define the maximal extractable randomness as

$$\ell^\varepsilon(X|B)_\rho := \max\{\ell \in \mathbb{N} \mid \exists \mathcal{P}: \log |\mathcal{Z}| \geq \ell \wedge d_{\text{sec}} \leq \varepsilon\},$$

where \mathcal{P} is any protocol defined above and $d_{\text{sec}}(\mathcal{P}, \rho_{XB}) := \min_{\sigma_B} P(\tau_{ZBS}, \pi_Z \otimes \sigma_B \otimes \tau_S)$ evaluated for the final state τ_{ZBS} obtained using \mathcal{P} and $\pi_Z = 1_Z/|\mathcal{Z}|$ fully mixed.

Note that the independence of the resulting randomness is often quantified using the trace distance instead of the purified distance. In particular, universal composability requires that

$$d_{\text{comp}}(\mathcal{P}, \rho_{XB}) := \frac{1}{2} \text{tr} |\tau_{XBS} - \pi_Z \otimes \tau_{BS}| \leq 2 d_{\text{sec}}(\mathcal{P}, \rho_{XB})$$

is small. However, we are only able to derive tight second order asymptotics for the relaxed requirement used to define ℓ^ε .

We show that the operational quantity ℓ^ε can be characterized by the smooth min-entropy in the following sense.

Theorem 8: Let $\rho_{XB} \in \mathcal{S}(\mathcal{H}_{XB})$ be a CQ state and $0 < \eta \leq \varepsilon < 1$. Then,

$$H_{\min}^{\varepsilon-\eta}(X|B)_\rho - \log \frac{1}{\eta^4} - 3 \leq \ell^\varepsilon(X|B)_\rho \leq H_{\min}^\varepsilon(X|B)_\rho.$$

We prove the direct part (left-hand inequality) and converse part (right-hand inequality) separately.

A. Proof of Converse

The converse employs the monotonicity of the smooth min-entropy under classical functions and is adapted from [19].

Proof of Converse of Theorem 8: We prove the statement by contradiction, i.e. we show that for all protocols \mathcal{P} ,

$$\log |\mathcal{Z}| > H_{\min}^\varepsilon(X|B)_\rho \implies d_{\text{sec}}(\mathcal{P}, \rho_{XB}) > \varepsilon, \quad (6)$$

which implies the converse.

For this purpose, let \mathcal{P} be any such protocol with $\log |\mathcal{Z}| > H_{\min}^\varepsilon(X|B)_\rho$. Hence, for any $s \in \mathcal{S}$, we have $H_{\min}^\varepsilon(Z|B)_{\tau^s} < \log |\mathcal{Z}|$ due to Proposition 3. By the definition of the smooth min-entropy, this implies that $\pi_Z \otimes \sigma_B \notin \mathcal{B}^\varepsilon(\tau_{ZB}^s)$ and, thus, $F(\tau_{ZB}^s, \pi_Z \otimes \sigma_B) < \sqrt{1 - \varepsilon^2}$ for any σ_B . Hence,

$$\begin{aligned} \max_{\sigma} F(\tau_{ZBS}, \pi_Z \otimes \sigma_B \otimes \tau_S) \\ = \max_{\sigma} \sum_s p_s F(\tau_{ZB}^s, \pi_Z \otimes \sigma_B) < \sqrt{1 - \varepsilon^2}, \end{aligned}$$

which shows the implication (6) and concludes the proof. ■

B. Proof of Achievability

We employ a result for two-universal hashing in [10].

Proof of Direct Part of Theorem 8: We employ Corollary 5.5.2 in [10], which states that, for any protocol \mathcal{P} using two-universal hashing, and any CQ state $\hat{\rho}_{XB} \in \mathcal{S}(\mathcal{H}_{XB})$, the state $\hat{\tau}_{ZBS} = \mathcal{F}(\hat{\rho}_{XB})$ satisfies

$$\text{tr} |\hat{\tau}_{ZBS} - \pi_Z \otimes \hat{\tau}_{BS}| \leq \sqrt{|\mathcal{Z}| 2^{-H_{\min}(X|B)_{\hat{\rho}}}}. \quad (7)$$

Now, let $\hat{\rho}_{XB} \in \mathcal{B}^{\varepsilon-\eta}(\rho_{XB})$ be such that $H_{\min}^{\varepsilon-\eta}(X|B)_\rho = H_{\min}(X|B)_{\hat{\rho}}$. Then,

$$\begin{aligned} d_{\text{sec}}(\mathcal{P}, \rho_{XB}) &\leq P(\tau_{ZBS}, \pi_Z \otimes \hat{\tau}_E \otimes \tau_S) \\ &\leq P(\hat{\tau}_{ZBS}, \pi_Z \otimes \hat{\tau}_E \otimes \tau_S) + P(\hat{\tau}_{ZBS}, \tau_{ZBS}) \\ &\leq \sqrt{2 \text{tr} |\hat{\tau}_{ZBS} - \pi_Z \otimes \hat{\tau}_E \otimes \tau_S|} + \varepsilon - \eta \\ &\leq \sqrt{2} (|\mathcal{Z}| 2^{-H_{\min}(X|B)_{\hat{\rho}}})^{\frac{1}{4}} + \varepsilon - \eta. \end{aligned}$$

Here, we used the Fuchs-van de Graaf inequality, Eq. (7), and the fact that $P(\hat{\tau}_{ZBS}, \tau_{ZBS}) \leq P(\hat{\rho}_{XB}, \rho_{XB}) \leq \varepsilon - \eta$. Thus, the condition $d_{\text{sec}}(\mathcal{P}, \rho_{XB}) \leq \varepsilon$ is satisfied when

$$\sqrt{2} (|\mathcal{Z}| 2^{-H_{\min}^{\varepsilon-\eta}(X|B)_\rho})^{\frac{1}{4}} \leq \eta$$

and we find that the choice

$$\ell = \log |\mathcal{Z}| = \lfloor H_{\min}^{\varepsilon-\eta}(X|B)_\rho + \log \eta^4 - 2 \rfloor$$

achieves the required bound. ■

IV. ONE-SHOT CHARACTERIZATION OF SOURCE COMPRESSION

Given a state $\rho_{XB} = \sum_x p_x |x\rangle\langle x| \otimes \phi_B^x$ that is classical on X , we are interested in compressing X into a register M such that M and the quantum system B allow to recover X with an error at most ε , where $0 < \varepsilon < 1$ is a fixed constant. We consider a general class of randomized compression protocols.

A protocol $\mathcal{P} = \{\mathcal{S}, \mathcal{M}, \{p_s\}, \{e_s\}, \{\mathcal{D}_{s,m}\}\}$ consists of a seed, \mathcal{S} , a code book, \mathcal{M} , and, for all $s \in \mathcal{S}$, a probability p_s and an encoder function $e_s : X \rightarrow M$. Furthermore, for every $s \in \mathcal{S}$ and every $m \in \mathcal{M}$, we have a decoder POVM $\mathcal{D}_{s,m} = \{M_{x,s,m}^s\}_x$ that measures X' on B . The protocol can be split into an encoding and decoding part:

- The encoding protocol acts on the initial state $\rho_{XBSS'} = \rho_{XB} \otimes \sum_s p_s |ss\rangle\langle ss|_{SS'}$ with the isometry $U_e = \sum_{x,s} |x\rangle\langle x|_X \otimes |e_s(x)\rangle_M \otimes |s\rangle\langle s|_S$. Informally, this means that, depending on the value $s \in \mathcal{S}$ in the register S , the function e_s is applied on X and the result is stored in M , while a copy of X is kept. This operation results in the state $\tau_{XMBSS'} := U_e \rho_{XBSS'} U_e^\dagger$.
- The decoding protocol acts on the systems S' , M and B to extract X' . We describe this operation using a TPCPM \mathcal{D} , which reads out $s \in \mathcal{S}$ from S' and $m \in \mathcal{M}$ from M and then applies the measurement $\mathcal{D}_{s,m}$ on B . This is given by the TP-CPM \mathcal{D} from MBS' to $X'S'$,

$$\mathcal{D} : \tau_{XMBSS'} \mapsto \sum_{x,s} |x\rangle\langle x|_{X'} \otimes |s\rangle\langle s|_{S'} \otimes$$

$$\text{tr}_{MBS'} (|s\rangle\langle s|_{S'} \otimes |m\rangle\langle m|_M \otimes M_{x,s,m}^s) \tau_{XMBSS'}.$$

We denote the final state by $\omega_{XX'SS'} := \mathcal{D}(\tau_{XMBSS'})$.

Using this class of protocols, we characterize the compressibility of a state ρ_{XB} by the minimal compression length (in bits), m^ε , required to achieve an error probability of at most ε .

Definition 6: Let $\rho_{XB} \in \mathcal{S}(\mathcal{H}_{XB})$ be a CQ state. Then, we define the minimal compression length as

$$m^\varepsilon(X|B)_\rho := \min\{m \in \mathbb{N} \mid \exists \mathcal{P}: \log |\mathcal{M}| \leq m \wedge p_{\text{err}} \leq \varepsilon\},$$

where \mathcal{P} is any protocol defined above and $p_{\text{err}}(\mathcal{P}, \rho_{XB}) := \mathbb{E}_{s \leftarrow p_s}(\Pr[X \neq X'|S = s])$ evaluated for the state $\omega_{XX'SS'}$ as defined above.⁴

We show that the operational quantity m^ε can be characterized by the hypothesis testing entropy in the following sense.

Theorem 9: Let $\rho_{XB} \in \mathcal{S}(\mathcal{H}_{XB})$ be a CQ state and $0 < \eta \leq \varepsilon < 1$. Then,

$$H_h^\varepsilon(X|B)_\rho \leq m^\varepsilon(X|B)_\rho \leq H_h^{\varepsilon-\eta}(X|B)_{\rho|\rho} + \log \frac{\varepsilon}{\eta^2} + 3.$$

We prove the direct part (right-hand inequality) and converse part (left-hand inequality) separately.

A. Proof of Converse

The proof of the converse utilizes various monotonicity properties of the hypothesis testing entropy.

Proof of Converse of Theorem 9: It is sufficient to show that, for every protocol \mathcal{P} as defined above, the requirement $p_{\text{err}}(\mathcal{P}, \rho_{XB}) \leq \varepsilon$ implies $H_h^\varepsilon(X|B)_\rho \leq \log |\mathcal{M}|$.

Let \mathcal{P} be any fixed protocol with $p_{\text{err}}(\mathcal{P}, \rho_{XB}) \leq \varepsilon$ and let $\rho_{XBSS'}$, $\tau_{XMBSS'}$ and $\omega_{XX'SS'}$ be defined as above. We employ the projector $Q_{XX'SS'} = \sum_{x,s} |xx\rangle\langle xx| \otimes |ss\rangle\langle ss|$ and note that $\text{tr}(Q_{XX'SS'} \omega_{XX'SS'}) = \sum_s p_s \Pr[X = X'|S = s] \geq 1 - \varepsilon$ due to the requirement on $p_{\text{err}}(\mathcal{P}, \rho_{XB})$. Thus, $Q_{XX'SS'}$ is primal feasible for $H_h^\varepsilon(XS|X'S')_\omega$ and we have

$$\begin{aligned} 0 &= \log \|Q_{X'S'}\| \\ &\geq H_h^\varepsilon(XS|X'S')_\omega \\ &\geq H_h^\varepsilon(XS|MBS')_\tau \quad (\text{cf. Prop. 4}) \\ &\geq H_h^\varepsilon(XMS|BS')_\tau - \log |\mathcal{M}| \quad (\text{cf. Prop. 7}) \\ &= H_h^\varepsilon(XS|BS')_\rho - \log |\mathcal{M}| \quad (\text{cf. Prop. 4}) \\ &= H_h^\varepsilon(X|B)_\rho - \log |\mathcal{M}|. \end{aligned}$$

The final equality follows from the following observation. Let Q_{AB} and $\{N_{AB}, \sigma_B, \eta\}$ be primal and dual optimal for $H_h^\varepsilon(X|B)_\rho$, respectively. Then, it is easy to verify that $Q_{AB} \otimes \sum_s |ss\rangle\langle ss|$ is primal feasible and $\{N_{AB} \otimes \rho_{SS'}, \sigma_B \otimes \rho_S, \eta\}$ is dual feasible for $H_h^\varepsilon(XS|BS')_\rho$, which implies equality. ■

B. Proof of Achievability

The proof is heavily based on [11] (see also [30], [20]); in particular, we need the following result [11]:

Lemma 10: For any $c > 0$, $0 \leq S \leq 1$ and $T \geq 0$, we have $1 - (S+T)^{-\frac{1}{2}} S(S+T)^{-\frac{1}{2}} \leq (1+c)(1-S) + (2+c+\frac{1}{c})T$.

The proof now employs two-universal hashing in the encoder as well as pretty good measurements in the decoder.

Proof of Direct Part of Theorem 9: We propose the following protocol. The encoder creates M from X through two-universal hashing [36], i.e. we consider a family of encoders and a seed satisfying $\Pr_{s \leftarrow p_s}[e_s(x) = e_s(z)] \leq \frac{1}{|\mathcal{M}|}$ if $x \neq z$. Let $Q_{XB} = \sum_x |x\rangle\langle x| \otimes Q_B^x$ be primal optimal for $H_h^{\varepsilon-\eta}(X|B)_{\rho|\rho}$. Then, we use the decoding POVMs

$\mathcal{D}_{s,m} = \{M_x^{s,m}\}_x$, where

$$M_x^{s,m} = \delta_{e_s(x),m} \left(\sum_{z \in e_s^{-1}(m)} Q_B^z \right)^{-\frac{1}{2}} Q_B^x \left(\sum_{z \in e_s^{-1}(m)} Q_B^z \right)^{-\frac{1}{2}}.$$

It remains to bound the average error of this protocol on the state ρ_{XB} . To do this, first note that $p_{\text{err}}(\mathcal{P}, \rho_{XB}) = \mathbb{E}_{s \leftarrow p_s, x \leftarrow p_x} [\text{tr}(\rho_B^x (1_B - M_x^{s,e_s(x)}))]$, where $1_B - M_x^{s,e_s(x)}$ can be upper-bounded using Lemma 10 as

$$\begin{aligned} 1_B - \left(\sum_{z, e_s(z)=e_s(x)} Q_B^z \right)^{-\frac{1}{2}} Q_B^x \left(\sum_{z, e_s(z)=e_s(x)} Q_B^z \right)^{-\frac{1}{2}} \\ \leq (1+c)(1_B - Q_B^x) + (2+c+\frac{1}{c}) \sum_{z \neq x} \delta_{e_s(z)=e_s(x)} Q_B^z \end{aligned}$$

for any $c > 0$ (to be optimized over below). We now substitute this to bound $p_{\text{err}}(\mathcal{P}, \rho_{XB})$, i.e.

$$\begin{aligned} p_{\text{err}}(\mathcal{P}, \rho_{XB}) &\leq (1+c) \mathbb{E}_{x \leftarrow p_x} [\text{tr}(\rho_B^x (1_B - Q_B^x))] \\ &\quad + (2+c+\frac{1}{c}) \mathbb{E}_{s \leftarrow p_s, x \leftarrow p_x} \left[\sum_{z \neq x} \delta_{e_s(z)=e_s(x)} \text{tr}(\rho_B^x Q_B^z) \right] \end{aligned}$$

The second expectation value can be simplified using the two-universal property: $\mathbb{E}_{s \leftarrow p_s} [\delta_{e_s(z)=e_s(x)}] \leq \frac{1}{|\mathcal{M}|}$ if $x \neq z$. We can thus upper bound the whole expression by

$$\begin{aligned} p_{\text{err}}(\mathcal{P}, \rho_{XB}) &\leq (1+c) \text{tr}(\rho_{BX} (1_{XB} - Q_{XB})) \\ &\quad + \frac{(c+1)^2}{c|\mathcal{M}|} \mathbb{E}_{x \leftarrow p_x} [\text{tr}(\rho_B^x \sum_{z \neq x} Q_B^z)] \\ &\leq (1+c)(\varepsilon - \eta) + \frac{(c+1)^2}{c|\mathcal{M}|} 2^{H_h^{\varepsilon-\eta}(X|B)_\rho}. \end{aligned}$$

Here, we used that $\text{tr}(\rho_{BX} Q_{XB}) \geq 1 - (\varepsilon - \eta)$ and

$$2^{H_h^{\varepsilon-\eta}(X|B)_{\rho|\rho}} = \text{tr}(\rho_B Q_B) \geq \sum_x p_x \text{tr}(\rho_B^x \sum_{z \neq x} Q_B^z).$$

Hence, this protocol will lead to an error of at most ε if we choose $m = \log |\mathcal{M}| = \left\lceil H_h^{\varepsilon-\eta}(X|B)_{\rho|\rho} + \log \frac{2+c+\frac{1}{c}}{\eta-c(\varepsilon-\eta)} \right\rceil$. The choice $c = \frac{\eta}{2\varepsilon-\eta}$ then leads to the desired bound. ■

V. HIERARCHY OF RELATIVE ENTROPIES

We will use the following properties of the relative entropies, which can be verified by a close inspection of their respective definitions.

- Monotonicity: For any TP-CPM \mathcal{E} ,

$$x = h, \max : D_x^\varepsilon(\rho||\sigma) \geq D_x^\varepsilon(\mathcal{E}(\rho)||\mathcal{E}(\sigma)). \quad (8)$$

- When $\sigma \leq \sigma'$, we find

$$x = h, \max : D_x^\varepsilon(\rho||\sigma) \geq D_x^\varepsilon(\rho||\sigma'). \quad (9)$$

Furthermore, if σ and σ' commute, this extends to $x = s$.

- When $\sigma' = 2^{-\lambda} \sigma$, we further have

$$x = h, \max, s : D_x^\varepsilon(\rho||\sigma) = D_x^\varepsilon(\rho||\sigma) + \lambda. \quad (10)$$

Furthermore, Lemma 9 of [28] is of pivotal for our analysis.

Lemma 11: For any $\rho, \sigma \in \mathcal{P}(\mathcal{H})$, we have $\rho \leq v(\sigma) \mathcal{E}_\sigma(\rho)$, where $v(\sigma)$ denotes the number of different eigenvalues of σ and \mathcal{E}_σ is a pinching that projects on the eigenspaces corresponding to the different eigenvalues of σ .

⁴Note that this definition implies that the optimal protocol for a fixed state ρ_{XB} is deterministic since we can always fix the seed to the value that achieves the lowest error probability in the average.

A. The Information Spectrum

We introduce the following quantity, which is as an entropic version of the quantum information spectrum [12]. (The relation of this quantity to the more traditional formulation of the information spectrum is explained in Section VIII.)

Definition 7: Let $\rho \in \mathcal{S}_{\leq}(\mathcal{H})$, $\sigma \in \mathcal{P}(\mathcal{H})$, and $0 \leq \varepsilon \leq 1$. Then, the information spectrum relative entropy is defined as

$$D_s^\varepsilon(\rho\|\sigma) := \sup \{ R \in \mathbb{R} \mid \text{tr } \rho\{\rho \leq 2^R \sigma\} \leq \varepsilon \}.$$

If ρ and σ commute, we may expand them in a common orthonormal eigenbasis, e.g. $\rho = \sum_y r_y |u_y\rangle\langle u_y|$ and $\sigma = \sum_y s_y |u_y\rangle\langle u_y|$. Consider now the distributions $P(y) = r_y$ and $Q(y) = s_y$, we find that $\text{tr } \rho\{\rho \leq e^R \sigma\} = P\{\log P - \log Q \leq R\}$, and recover the classical information spectrum $D_s^\varepsilon(P\|Q)$ as defined in (2).

The information spectrum is intimately related to hypothesis testing, as has been pointed out in [12]. Here, we present a proof in the one-shot setting for the convenience of the reader.

Lemma 12: Let $\rho \in \mathcal{S}(\mathcal{H})$, $\sigma \in \mathcal{P}(\mathcal{H})$, and $\delta > 0$. Then,

$$D_s^\varepsilon(\rho\|\sigma) \leq D_h^\varepsilon(\rho\|\sigma) \leq D_s^{\varepsilon+\delta}(\rho\|\sigma) - \log \delta. \quad (11)$$

Proof: The first inequality follows by considering the projector $Q = \{\rho > 2^R \sigma\}$ that is primal feasible for $D_h^\varepsilon(\rho\|\sigma)$ when $R = D_s^\varepsilon(\rho\|\sigma) - \xi$ for an arbitrary $\xi > 0$. Furthermore,

$$\langle \sigma, Q \rangle = \text{tr } \sigma\{\rho > 2^R \sigma\} \leq 2^{-R} \text{tr } \rho\{\rho > 2^R \sigma\} \leq 2^{-R}.$$

Hence, $D_h^\varepsilon(\rho\|\sigma) \geq R$, which implies the result when $\xi \rightarrow 0$.

To get the second inequality, consider first the case where $D_h^\varepsilon(\rho\|\sigma)$ is finite, and an operator $0 \leq Q \leq 1$ that is primal optimal for $D_h^\varepsilon(\rho\|\sigma)$. Using $\mu = \log \delta + D_h^\varepsilon(\rho\|\sigma)$, we find

$$\begin{aligned} \text{tr } \rho\{\rho > 2^\mu \sigma\} &\geq \text{tr } ((\rho - 2^\mu \sigma)\{\rho > 2^\mu \sigma\}) \\ &\geq \text{tr } ((\rho - 2^\mu \sigma)Q) = \langle \rho, Q \rangle - 2^\mu \langle \sigma, Q \rangle \\ &\geq 1 - \varepsilon - \delta, \end{aligned} \quad (12)$$

where the last inequality follows from the fact that Q is primal optimal. Thus, $D_s^{\varepsilon+\delta}(\rho\|\sigma) \geq \mu$, concluding the proof. Finally, in the case where $\langle \sigma, Q \rangle = 0$, Eq. (12) holds for any μ and, thus, both sides of the inequality diverge. ■

Furthermore, we consider the information spectrum for the state $\mathcal{E}_\sigma(\rho)$, where \mathcal{E}_σ is a pinching of ρ in the eigenbasis of σ , i.e.

$$\mathcal{E}_\sigma(\rho) = \sum_s P_\sigma^s \rho P_\sigma^s, \quad \text{where} \quad P_\sigma^s = \sum_{y:s_y=s} |u_y\rangle\langle u_y|$$

is the projector onto the eigenspace with eigenvalue s . We will see in the following that the entropies $D_s^\varepsilon(\mathcal{E}_\sigma(\rho)\|\sigma)$ and $D_s^\varepsilon(\rho\|\sigma)$ are related. Furthermore, $\mathcal{E}_\sigma(\rho)$ and σ commute.

In order to refine this analysis and make it applicable for the second order expansion, we employ the probability introduced by Nussbaum and Szkoła [29]. Using the eigenvalue decompositions $\rho = \sum_x r_x |v_x\rangle\langle v_x|$ and $\sigma = \sum_y s_y |u_y\rangle\langle u_y|$, they defined two distributions:

$$P_{\rho,\sigma}(x, y) := r_x |\langle v_x | u_y \rangle|^2 \quad \text{and} \quad Q_{\rho,\sigma}(x, y) := s_y |\langle v_x | u_y \rangle|^2.$$

These distributions have the very convenient property that the first two moments of $\log P_{\rho,\sigma} - \log Q_{\rho,\sigma}$ under $P_{\rho,\sigma}$ agree

with the respective moments of $\log \rho - \log \sigma$ under ρ . Namely, it is easy to verify that

$$D(P_{\rho,\sigma}\|Q_{\rho,\sigma}) = D(\rho\|\sigma) \quad \text{and} \quad s(P_{\rho,\sigma}\|Q_{\rho,\sigma}) = s(\rho\|\sigma). \quad (13)$$

Moreover, in the i.i.d. scenario, we have $P_{\rho^n,\sigma^n} = P_{\rho,\sigma}^n$ and $Q_{\rho^n,\sigma^n} = Q_{\rho,\sigma}^n$ using the notation introduced previously.

The asymptotic analysis can thus be reduced to the problem of finding suitable relations between the one-shot entropies and the quantity $D_s^\varepsilon(P_{\rho,\sigma}\|Q_{\rho,\sigma})$ for general ρ and σ .

B. Useful Inequalities for Relative Entropies

We obtain the following inequalities for relative entropies.

Proposition 13: Let $\rho \in \mathcal{S}(\mathcal{H})$, $\sigma \in \mathcal{P}(\mathcal{H})$, $0 < \varepsilon < 1$, and $0 < \delta < \varepsilon$. Then, using $\nu = \nu(\sigma)$, we obtain

$$D_{\max}^{\sqrt{1-\varepsilon}}(\rho\|\sigma) \geq D_s^{\varepsilon-\delta}(\rho\|\sigma) + 2 \log \delta - 2 - \log \varepsilon, \quad (14)$$

$$D_{\max}^{\sqrt{1-\varepsilon}}(\rho\|\sigma) \leq D_s^\varepsilon(\mathcal{E}_\sigma(\rho)\|\sigma) + \log \nu - \log(1 - \varepsilon), \quad (15)$$

$$D_s^{\varepsilon-\delta}(\mathcal{E}_\sigma(\rho)\|\sigma) \leq D_s^\varepsilon(P_{\rho,\sigma}\|Q_{\rho,\sigma}) - \log \delta, \quad (16)$$

$$D_s^\varepsilon(\mathcal{E}_\sigma(\rho)\|\sigma) \geq D_s^{\varepsilon-\delta}(P_{\rho,\sigma}\|Q_{\rho,\sigma}) + \log \delta - \log \nu, \quad (17)$$

$$D_{\max}^{\sqrt{1-\varepsilon}}(\rho\|\sigma) \leq D_h^\varepsilon(\rho\|\sigma) + \log \nu - \log(1 - \varepsilon), \quad (18)$$

$$D_{\max}^{\sqrt{1-\varepsilon}}(\rho\|\sigma) \geq D_h^{\varepsilon-\delta}(\rho\|\sigma) + 3 \log \delta - 3 \log 3 - \log \varepsilon. \quad (19)$$

Proof of (14): Assume $D_{\max}^{\sqrt{1-\varepsilon}}(\rho\|\sigma) = R$ and choose $\tilde{\rho}$ such that $\tilde{\rho} \leq 2^R \sigma$ and $F(\tilde{\rho}, \rho)^2 \geq \varepsilon$. Now, we consider the binary projective measurement $\{\{\rho \leq 2^{R+\delta'} \sigma\}, \{\rho > 2^{R+\delta'} \sigma\}\}$ for some $\delta' \geq 0$. The monotonicity of F yields

$$\begin{aligned} \sqrt{\varepsilon} &\leq \sqrt{\text{tr } \rho\{\rho \leq 2^{R+\delta'} \sigma\}} \sqrt{\text{tr } \tilde{\rho}\{\rho \leq 2^{R+\delta'} \sigma\}} \\ &\quad + \sqrt{\text{tr } \rho\{\rho > 2^{R+\delta'} \sigma\}} \sqrt{\text{tr } \tilde{\rho}\{\rho > 2^{R+\delta'} \sigma\}} \\ &\leq \sqrt{\text{tr } \rho\{\rho \leq 2^{R+\delta'} \sigma\}} + \sqrt{\text{tr } \tilde{\rho}\{\rho > 2^{R+\delta'} \sigma\}}. \end{aligned} \quad (20)$$

Moreover, the condition $\tilde{\rho} \leq 2^R \sigma$ implies that

$$\begin{aligned} \text{tr } \tilde{\rho}\{\rho > 2^{R+\delta'} \sigma\} &\leq 2^R \text{tr } \sigma\{\rho > 2^{R+\delta'} \sigma\} \\ &\leq 2^{-\delta'} \text{tr } \rho\{\rho > 2^{R+\delta'} \sigma\} \leq 2^{-\delta'}. \end{aligned}$$

Combining this with (20) and choosing $\sqrt{2^{-\delta'}} = \sqrt{\varepsilon} - \sqrt{\varepsilon - \delta}$, we find $\text{tr } \rho\{\rho \leq 2^{R+\delta'} \sigma\} \geq \varepsilon - \delta$. Hence,

$$D_s^{\varepsilon-\delta}(\rho\|\sigma) < R + \delta' \leq R + \log \frac{4\varepsilon}{\delta^2},$$

where we used $\sqrt{\varepsilon} - \sqrt{\varepsilon - \delta} \geq \frac{\delta}{2\sqrt{\varepsilon}}$ in the last step. ■

Proof of (15): We set $R = D_s^\varepsilon(\mathcal{E}_\sigma(\rho)\|\sigma)$ and define $Q = \{\mathcal{E}_\sigma(\rho) \leq 2^R \sigma\}$. Thus, Q satisfies $\text{tr}(\rho Q) = \text{tr}(\mathcal{E}_\sigma(\rho)Q) = \varepsilon$, where we used that \mathcal{E}_σ leaves Q invariant. Now, we choose

$$\tilde{\rho} = \frac{Q \rho Q}{\text{tr}(\rho Q)} \quad \text{s.t.} \quad F(\tilde{\rho}, \rho)^2 \geq \text{tr}(\rho Q) = \varepsilon.$$

Moreover, Lemma 11 shows that $\rho \leq \nu \mathcal{E}_\sigma(\rho)$ and, thus,

$$Q \rho Q \leq \nu Q \mathcal{E}_\sigma(\rho) Q \leq \nu 2^R Q \sigma Q \leq \nu 2^R \sigma.$$

where we used the definition of Q and that it commutes with σ in the final two inequalities. Thus, $\tilde{\rho} \leq \frac{\nu 2^R}{1-\varepsilon} \sigma$ and

$$D_{\max}^{\sqrt{1-\varepsilon}}(\rho\|\sigma) \leq \inf \{ \lambda \mid \tilde{\rho} \leq 2^\lambda \sigma \} \leq \log \nu + R - \log(1 - \varepsilon),$$

completing the proof. ■

Proof of (16) and (17): Since $\mathcal{E}_\sigma(\rho)$ commutes with σ , there exists a common eigenvector system $\{u_y\}$, i.e.

$$\mathcal{E}_\sigma(\rho) = \sum_y r'_y |u_y\rangle\langle u_y| \quad \text{and} \quad \sigma = \sum_y s_y |u_y\rangle\langle u_y|.$$

Using the representation $\rho = \sum_x r_x |v_x\rangle\langle v_x|$, we can describe distributions $P_{\rho,\sigma}$ and $Q_{\rho,\sigma}$ as follows

$$P_{\rho,\sigma}(x, y) = r_x |\langle v_x | u_y \rangle|^2 \quad \text{and} \quad Q_{\rho,\sigma}(x, y) = s_y |\langle v_x | u_y \rangle|^2.$$

Furthermore, we define the distribution

$$Q'_{\rho,\sigma}(x, y) := r'_y |\langle v_x | u_y \rangle|^2$$

and note that $D_s^\varepsilon(\mathcal{E}_\sigma(\rho) \| \sigma) = D_s^\varepsilon(Q'_{\rho,\sigma} \| Q_{\rho,\sigma})$. We drop the subscripts ρ and σ in the following. For real R and δ' , we find

$$\begin{aligned} & P\left\{\log \frac{P}{Q} \leq R\right\} \\ &= P\left\{\log \frac{Q'}{Q} + \log \frac{P}{Q'} \leq R\right\} \geq \\ &\geq P\left\{\log \frac{Q'}{Q} \leq R - \delta' \wedge \log \frac{P}{Q'} \leq \delta'\right\} \\ &= 1 - P\left\{\log \frac{Q'}{Q} > R - \delta' \vee \log \frac{P}{Q'} > \delta'\right\} \\ &\geq 1 - P\left\{\log \frac{Q'}{Q} > R - \delta'\right\} - P\left\{\log \frac{P}{Q'} > \delta'\right\} \\ &= P\left\{\log \frac{Q'}{Q} \leq R - \delta'\right\} - P\left\{\log \frac{P}{Q'} > \delta'\right\}. \end{aligned} \quad (21)$$

Similarly, we bound

$$\begin{aligned} & P\left\{\log \frac{Q'}{Q} \leq R + \delta'\right\} \geq \\ & P\left\{\log \frac{P}{Q} \leq R\right\} - P\left\{\log \frac{P}{Q'} < -\delta'\right\}. \end{aligned} \quad (22)$$

Moreover, we have $P\{\log \frac{P}{Q'} < -\delta'\} \leq P\{P < \delta\} < \delta$, where we chose $\delta' = -\log \delta$. Hence, if we further choose $R = D_s^\varepsilon(P \| Q)$, we have $P\{\log \frac{P}{Q} \leq R\} \leq \varepsilon$ by definition. Together with (22), this yields

$$P\left\{\log \frac{Q'}{Q} \leq R + \delta'\right\} > \varepsilon - \delta,$$

which directly implies (16).

Applying the Markov inequality, we obtain

$$\begin{aligned} P\left\{\log \frac{P}{Q'} > \delta'\right\} &= P\left\{\frac{P}{Q'} > 2^{-\delta'}\right\} \\ &\leq \sum_{x,y} P(x, y) \frac{P(x, y)}{Q'(x, y)} 2^{-\delta'} \\ &= 2^{-\delta'} \text{tr}(\rho^2 \mathcal{E}_\sigma(\rho)^{-1}). \end{aligned}$$

Since the quantity $\text{tr}(\rho^2 \sigma^{-1})$ decreases under the operation of TP-CPMs [37], it also satisfies joint convexity. Hence, using the eigenvalue decomposition of ρ , we have

$$\begin{aligned} \text{tr}(\rho^2 \mathcal{E}_\sigma(\rho)^{-1}) &\leq \sum_x r_x \text{tr}(|v_x\rangle\langle v_x|^2 \mathcal{E}_\sigma(|v_x\rangle\langle v_x|)^{-1}) \\ &\leq \max_\phi \langle \phi | \mathcal{E}_\sigma(|\phi\rangle\langle \phi|)^{-1} | \phi \rangle. \end{aligned} \quad (23)$$

Furthermore, since \mathcal{E}_σ is a projective measurement of the form $\{M_i\}_{i=1}^\nu$, we may write $\mathcal{E}_\sigma(|\phi\rangle\langle \phi|) = \sum_{i=1}^\nu \alpha_i |\psi_i\rangle\langle \psi_i|$ for coefficients $\alpha_i = \langle \phi | M_i | \phi \rangle$ and orthonormal vectors $|\psi_i\rangle = \frac{1}{\sqrt{\alpha_i}} M_i |\phi\rangle$. The expression (23) now yields

$$\langle \phi | \mathcal{E}_\sigma(|\phi\rangle\langle \phi|)^{-1} | \phi \rangle = \sum_{\substack{i=1 \\ \alpha_i > 0}}^\nu \frac{1}{\alpha_i} |\langle \phi | \psi_i \rangle|^2 = \sum_{\substack{i=1 \\ \alpha_i > 0}}^\nu 1 \leq \nu.$$

Finally, we thus find $P\{\log \frac{P}{Q'} > \delta'\} \leq 2^{-\delta'} \nu$. The choices $\delta = 2^{-\delta'} \nu$ and $R = D_{\varepsilon-\delta}(P \| Q)$ together with (21) yield

$$P\left\{\log \frac{Q'}{Q} \leq R - \log \frac{\nu}{\delta}\right\} \leq P\left\{\log \frac{P}{Q} \leq R\right\} + \delta \leq \varepsilon,$$

which concludes the proof. \blacksquare

Proof of (18) and (19): The last two inequalities follow from the previous relations. We have

$$D_{\max}^{\sqrt{1-\varepsilon}}(\rho \| \sigma) \leq D_s^\varepsilon(\mathcal{E}_\sigma(\rho) \| \sigma) + \log \nu - \log(1 - \varepsilon) \quad (\text{cf. (15)})$$

$$\leq D_h^\varepsilon(\mathcal{E}_\sigma(\rho) \| \sigma) + \log \nu - \log(1 - \varepsilon) \quad (\text{cf. (11)})$$

$$\leq D_h^\varepsilon(\rho \| \sigma) + \log \nu - \log(1 - \varepsilon). \quad (\text{cf. (8)})$$

Furthermore, using Eqs. (14) and (11)

$$\begin{aligned} D_{\max}^{\sqrt{1-\varepsilon}}(\rho \| \sigma) &\geq D_s^{\varepsilon-\delta_1}(\rho \| \sigma) + 2 \log \delta_1 - 2 - \log \varepsilon \\ &\geq D_h^{\varepsilon-\delta_1-\delta_2}(\rho \| \sigma) + \log \delta_1^2 \delta_2 - 2 - \log \varepsilon, \end{aligned}$$

and the choice $\delta_1 = \frac{2\delta}{3}$, $\delta_2 = \frac{\delta}{3}$ yields (19). \blacksquare

C. One-Shot Entropies and the Information Spectrum

The above relations allow us to bound the hypothesis testing and smooth entropies in terms of the classical information spectrum of $P_{\rho,\sigma}$ and $Q_{\rho,\sigma}$.

In order to refine these statements, we need the following notation. For a given positive semi-definite matrix σ , we denote the number of distinct eigenvalues of σ by $\nu(\sigma)$. We also define the number $\lambda(\sigma) := \log \lambda_{\max}(\sigma) - \log \lambda_{\min}(\sigma)$, where λ_{\max} is the maximum and λ_{\min} the minimum eigenvalue of σ . Finally, we employ

$$\theta(\sigma) := \min\{2\lceil \lambda(\sigma) \rceil, \nu(\sigma)\}.$$

The bounds can now be stated as follows.

Theorem 14: Let $\rho \in \mathcal{S}(\mathcal{H})$, $\sigma \in \mathcal{P}(\mathcal{H})$ and $0 < \varepsilon < 1$ and $0 < \delta < \min\{\varepsilon, 1 - \varepsilon\}$. Then, using $\theta = \theta(\sigma)$, $P = P_{\rho,\sigma}$ and $Q = Q_{\rho,\sigma}$, we have

$$D_h^\varepsilon(\rho \| \sigma) \leq D_s^{\varepsilon+\delta}(P \| Q) + \log \frac{2^8 \varepsilon \theta}{\delta^4 (1 - \varepsilon - \delta)}, \quad (24)$$

$$D_h^\varepsilon(\rho \| \sigma) \geq D_s^{\varepsilon-\delta}(P \| Q) - \log \theta + \log \delta, \quad (25)$$

$$D_{\max}^{\sqrt{1-\varepsilon}}(\rho \| \sigma) \leq D_s^{\varepsilon+\delta}(P \| Q) + \log \theta - \log(\delta(1 - \varepsilon)), \quad (26)$$

$$D_{\max}^{\sqrt{1-\varepsilon}}(\rho \| \sigma) \geq D_s^{\varepsilon-\delta}(P \| Q) - \log(3^3 \varepsilon \theta) + \log \delta^3. \quad (27)$$

Proof: We first show weaker inequalities for $\nu(\sigma)$ in place of θ and then argue that the inequalities still hold if we replace ν by $2\lceil \lambda(\sigma) \rceil$. In particular, this implies that they also holds for the minimum of these two expression, i.e. for θ .

Let $\nu = \nu(\sigma)$, and $\delta_i > 0, i = 1 \dots 3$, such that $\sum_i \delta_i = \delta$. (We will optimize over these partitions later.) We find

$$D_h^\varepsilon(\rho||\sigma) \leq D_s^{\varepsilon+\delta_1}(\rho||\sigma) - \log \delta_1 \quad (\text{cf. (11)})$$

$$\leq D_{\max}^{\sqrt{1-\varepsilon-\delta_1-\delta_2}}(\rho||\sigma) - \log \delta_1 \delta_2^2 + \log(4\varepsilon) \quad (\text{cf. (14)})$$

$$\leq D_s^{\varepsilon+\delta_1+\delta_2}(\mathcal{E}_\sigma(\rho)||\sigma) - \log \delta_1 \delta_2^2 + \log(4\varepsilon) - \log(1-\varepsilon-\delta_1-\delta_2) + \log \nu \quad (\text{cf. (15)})$$

$$\leq D_s^{\varepsilon+\delta_1+\delta_2+\delta_3}(P||Q) - \log \delta_1 \delta_2^2 \delta_3 + \log(4\varepsilon) - \log(1-\varepsilon-\delta_1-\delta_2) + \log \nu. \quad (\text{cf. (16)})$$

Choosing $\delta_1 = \delta_3 = \frac{\delta}{4}$, $\delta_2 = \frac{\delta}{2}$ yields (24). Next, we have

$$D_h^\varepsilon(\rho||\sigma) \geq D_h^\varepsilon(\mathcal{E}_\sigma(\rho)||\sigma) \quad (\text{cf. (8)})$$

$$\geq D_s^\varepsilon(\mathcal{E}_\sigma(\rho)||\sigma) \quad (\text{cf. (11)})$$

$$\geq D_s^{\varepsilon-\delta}(P||Q) + \log \delta - \log \nu, \quad (\text{cf. (17)})$$

which constitutes (25). Then, Eq. (26) follows from

$$D_{\max}^{\sqrt{1-\varepsilon}}(\rho||\sigma) \leq D_s^\varepsilon(\mathcal{E}_\sigma(\rho)||\sigma) + \log \nu - \log(1-\varepsilon) \quad (\text{cf. (15)})$$

$$\leq D_s^{\varepsilon+\delta}(P||Q) - \log \delta + \log \nu - \log(1-\varepsilon). \quad (\text{cf. (16)})$$

Finally we show (27). For any $\delta_1, \delta_2 > 0$ such that $\delta_1 + \delta_2 = \delta$,

$$D_{\max}^{\sqrt{1-\varepsilon}}(\rho||\sigma) \geq D_{\max}^{\sqrt{1-\varepsilon}}(\mathcal{E}_\sigma(\rho)||\sigma) \quad (\text{cf. (8)})$$

$$\geq D_s^{\varepsilon-\delta_1}(\mathcal{E}_\sigma(\rho)||\sigma) + 2 \log \delta_1 - \log(4\varepsilon) \quad (\text{cf. (14)})$$

$$\geq D_s^{\varepsilon-\delta_1-\delta_2}(P||Q) + \log \delta_1^2 \delta_2 - \log(4\varepsilon \nu). \quad (\text{cf. (17)})$$

Choosing $\delta_1 = \frac{2\delta}{3}$, $\delta_2 = \frac{\delta}{3}$, we obtain (27).

The above inequalities can now be adapted such that ν is replaced by $2\lceil \lambda(\sigma) \rceil$. We exemplify this by proving the inequality corresponding to (24). However, the argument is analogous for all inequalities in the theorem.

For a positive integer l to be determined later, we define σ' by the following procedure. First, we diagonalize $\sigma = \sum_y s_y |u_y\rangle\langle u_y|$ with $s_1 \geq s_2 \geq \dots \geq s_d$ and define $\lambda = \lambda(\sigma) = \log s_1 - \log s_d$. Moreover, we define $s'_y = s_d 2^{\frac{\lambda}{l}}$ when $\log s_y \in (\log s_d + \frac{\lambda}{l}i, \log s_d + \frac{\lambda}{l}(i+1)]$ for $i = 0 \dots l-1$ and $s'_y = s_d$ if $s_y = s_d$. Hence, $\sigma' \leq \sigma$ and $\sigma'' := 2^{-\frac{\lambda}{l}}\sigma \leq \sigma'$. Since the number of eigenvectors of σ' is at most l , (24) yields

$$D_h^\varepsilon(\rho||\sigma) \leq D_h^\varepsilon(\rho||\sigma') \quad (\text{cf. (9)})$$

$$\leq D_s^{\varepsilon+\delta}(P_{\rho,\sigma'}||Q_{\rho,\sigma'}) + \log \frac{2^8 \varepsilon l}{\delta^4(1-\varepsilon-\delta)} \quad (\text{cf. (24)})$$

$$\leq D_s^{\varepsilon+\delta}(P_{\rho,\sigma''}||Q_{\rho,\sigma''}) + \log \frac{2^8 \varepsilon l}{\delta^4(1-\varepsilon-\delta)} \quad (\text{cf. (9)})$$

$$= D_s^{\varepsilon+\delta}(P_{\rho,\sigma}||Q_{\rho,\sigma}) + \log \frac{2^8 \varepsilon l}{\delta^4(1-\varepsilon-\delta)} + \frac{\lambda}{l}. \quad (\text{cf. (10)})$$

Finally, substituting $\lceil \lambda(\sigma) \rceil$ into l , we find $\frac{\lambda}{l} + \log l \leq \log 2\lceil \lambda(\sigma) \rceil$ and, thus,

$$D_h^\varepsilon(\rho||\sigma) \leq D_s^{\varepsilon+\delta}(P_{\rho,\sigma}||Q_{\rho,\sigma}) + \log \frac{2^8 \varepsilon \cdot 2\lceil \lambda \rceil}{\delta^4(1-\varepsilon-\delta)}.$$

VI. ASYMPTOTIC ANALYSIS

We first investigate the behavior of the classical information spectrum of the Nussbaum-Szkoła distributions in the asymptotic limit. For this purpose, we consider the quantity

$$D_s^\varepsilon(P_{\rho,\sigma}||Q_{\rho,\sigma}) = \sup\{R \in \mathbb{R} \mid P\{Z \leq R\} \leq \varepsilon\},$$

which is equivalent to $F_Z^{-1}(\varepsilon)$, the inverse of the cumulative distribution function of $\bar{Z} = \log P_{\rho,\sigma}(X) - \log Q_{\rho,\sigma}(X)$.

In particular, we are interested in i.i.d. states $\rho^n = \rho^{\otimes n}$ and $\sigma^n = \sigma^{\otimes n}$. Then, the respective Nussbaum-Szkoła distributions are also of the i.i.d. form $P_{\rho,\sigma}^n(\vec{x}) = \prod_i P_{\rho,\sigma}(x_i)$ and, similarly, $Q_{\rho,\sigma}^n(\vec{x}) = \prod_i Q_{\rho,\sigma}(x_i)$. It is easy to verify that the information spectrum evaluates to

$$D_s^\varepsilon(P_{\rho,\sigma}^n||Q_{\rho,\sigma}^n) = n \sup\{R \mid P^n\{\bar{Z} \leq R\} \leq \varepsilon\}, \quad (28)$$

where $\bar{Z} = \frac{1}{n} \sum_i Z_i$ is averaged over n i.i.d. random variables $Z_i = \log P_{\rho,\sigma} - \log Q_{\rho,\sigma}$. Now, due to the central limit theorem, the distribution of

$$\sqrt{n} \frac{\bar{Z} - \mu}{s}, \quad \text{where } \mu = \mathbb{E}[Z] \text{ and } s = \sqrt{\mathbb{E}[(Z - \mu)^2]}$$

converges to the normal distribution. More precisely, the Berry-Esseen theorem states that

$$\left| P^n \left\{ \sqrt{n} \frac{\bar{Z} - \mu}{s} \leq z \right\} - \Phi(z) \right| \leq \frac{Cr^3}{s^3 \sqrt{n}},$$

where $r = \sqrt[3]{\mathbb{E}[|Z - \mu|^3]}$, $C < \frac{1}{2}$ [38], and the cumulative standard Gaussian distribution is given by

$$\Phi(x) := \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{x^2/2} dx.$$

We now evaluate the terms, using (13). We get

$$\mu = D(P_{\rho,\sigma}||Q_{\rho,\sigma}) = D(\rho||\sigma),$$

$$s = s(P_{\rho,\sigma}||Q_{\rho,\sigma}) = s(\rho||\sigma) \quad \text{and}$$

$$r = r(P_{\rho,\sigma}||Q_{\rho,\sigma}), \quad \text{where}$$

$$r(P_{\rho,\sigma}||Q_{\rho,\sigma}) := \sqrt[3]{\sum_x P_{\rho,\sigma}(x) |\log P_{\rho,\sigma}(x) - \log Q_{\rho,\sigma}(x) - \mu|^3}.$$

Combining this with (28), we can write

$$D_s^\varepsilon(P_{\rho,\sigma}^n||Q_{\rho,\sigma}^n) = nD(\rho||\sigma) + \sqrt{n} s(\rho||\sigma) \cdot \sup \left\{ x \mid P^n \left\{ \sqrt{n} \frac{\bar{Z} - \mu}{s} \leq x \right\} \leq \varepsilon \right\},$$

and further use the Berry-Esseen theorem to bound

$$D_s^\varepsilon(P_{\rho,\sigma}^n||Q_{\rho,\sigma}^n) \leq nD(\rho||\sigma) + \sqrt{n} s(\rho||\sigma) \Phi^{-1} \left(\varepsilon + \frac{Cr^3}{s^3 \sqrt{n}} \right) \\ D_s^\varepsilon(P_{\rho,\sigma}^n||Q_{\rho,\sigma}^n) \geq nD(\rho||\sigma) + \sqrt{n} s(\rho||\sigma) \Phi^{-1} \left(\varepsilon - \frac{Cr^3}{s^3 \sqrt{n}} \right). \quad (29)$$

Finally, assume that r is finite. Then, since Φ^{-1} is continuously differentiable, for any fixed $\varepsilon \in (0, 1)$ and δ proportional

to $1/\sqrt{n}$, we have the following asymptotic expansion for large n :⁵

$$D_s^{\varepsilon \pm \delta}(P_{\rho, \sigma}^n \| Q_{\rho, \sigma}^n) = nD(\rho \| \sigma) + \sqrt{n} s(\rho \| \sigma) \Phi^{-1}(\varepsilon) + O(1). \quad (30)$$

A. Asymptotic Behavior of Relative Entropies

We first investigate the asymptotic behavior of $D_h^\varepsilon(\rho^n \| \sigma^n)$ and $D_{\max}^\varepsilon(\rho^n \| \sigma^n)$ for large n . A straight-forward application of Theorem 14 yields, for $0 < \delta < \min\{\varepsilon, 1 - \varepsilon\}$,

$$\begin{aligned} D_s^{\varepsilon - \delta}(P_{\rho, \sigma}^n \| Q_{\rho, \sigma}^n) - \log \frac{\theta(\sigma^n)}{\delta} &\leq D_h^\varepsilon(\rho^n \| \sigma^n) \\ &\leq D_s^{\varepsilon + \delta}(P_{\rho, \sigma}^n \| Q_{\rho, \sigma}^n) + \log \frac{2^8 \varepsilon \theta(\sigma^n)}{\delta^4(1 - \varepsilon - \delta)}, \end{aligned}$$

Now, we observe that $\theta(\sigma_n) \leq 2\lceil \lambda(\sigma^n) \rceil = 2\lceil n\lambda(\sigma) \rceil$ scales at most linearly in n if $\lambda(\sigma)$ is finite. Furthermore, choosing $\delta = 1/\sqrt{n}$, we can apply (30) to get

$$D_h^\varepsilon(\rho^n \| \sigma^n) = nD(\rho \| \sigma) + \sqrt{n} s(\rho \| \sigma) \Phi^{-1}(\varepsilon) + O(\log n). \quad (31)$$

An analogous relation is derived for $D_{\max}^\varepsilon(\rho^n \| \sigma^n)$, where we use Theorem 14 and the relation $\Phi^{-1}(1 - \varepsilon^2) = -\Phi^{-1}(\varepsilon^2)$:

$$D_{\max}^\varepsilon(\rho^n \| \sigma^n) = nD(\rho \| \sigma) - \sqrt{n} s(\rho \| \sigma) \Phi^{-1}(\varepsilon^2) + O(\log n). \quad (32)$$

B. Asymptotic Behavior of Operational Quantities

We first treat source compression with quantum side information. Recall Theorem 9, which provides the following bounds on m^ε .

For any CQ state ρ_{XB} and $0 < \nu \leq \varepsilon < 1$, we have

$$\begin{aligned} -D_h^\varepsilon(\rho_{XB} \| 1_X \otimes \rho_B) &\leq \max_{\sigma_B} -D_h^\varepsilon(\rho_{XB} \| 1_X \otimes \sigma_B) \\ &\leq m^\varepsilon(X|B)_\rho \leq -D_h^{\varepsilon - \eta}(\rho_{XB} \| 1_X \otimes \rho_B) + \log \frac{2^3 \varepsilon}{\eta^2}. \end{aligned} \quad (33)$$

We now consider the i.i.d. asymptotic setting with ρ_{XB}^n and its marginal ρ_B^n . First, note that, $\lambda(\rho_B^n) = n\lambda(\rho_B)$ is linear in n , and, thus, $\log \theta(\rho_B^n)$ is of the order $O(\log n)$. Furthermore, the choice $\eta = 1/\sqrt{n}$ ensures that the additive terms in (33) are of the order $O(\log n)$.

Combined with (31), this yields the following result.

Corollary 15: For any CQ state ρ_{XB} and any $0 < \varepsilon < 1$, we find the following i.i.d. asymptotic expansion:

$$m^\varepsilon(X^n|B^n) = nH(X|B) - \sqrt{n} s(X|B) \Phi^{-1}(\varepsilon) + O(\log n).$$

To analyze source compression with quantum side information, we start with the one-shot characterization of ℓ^ε given in Theorem 8. For any CQ state ρ_{XB} and $0 < \varepsilon < 1$, we have

$$\begin{aligned} -D_{\max}^{\varepsilon - \eta}(\rho_{XB} \| 1_X \otimes \rho_B) - \log \frac{2^3}{\eta^4} &\leq \ell^\varepsilon(X|B)_\rho \\ &\leq \max_{\sigma_B} -D_{\max}^\varepsilon(\rho_{XB} \| 1_X \otimes \sigma_B). \end{aligned}$$

⁵If f is continuously differentiable, c a constant and $n \geq n_0$, we may write $\sqrt{n}f(x + \frac{c}{\sqrt{n}}) = \sqrt{n}f(x) + cf'(a)$ for some $a \in [x, x + \frac{c}{\sqrt{n_0}}]$.

Note that a simple application of Theorem 14 is not sufficient for deriving the i.i.d. asymptotic expansion of $\ell^\varepsilon(X|B)_\rho$ because of the optimization concerning σ_B and the fact that we cannot bound $\theta(\sigma_B)$ for the optimal σ_B .

Instead, we use the following relation (cf. Proposition 13),

$$\begin{aligned} D_{\max}^\varepsilon(\rho_{XB} \| 1_X \otimes \sigma_B) \\ \geq D_h^{1 - \varepsilon^2 - \mu}(\rho_{XB} \| 1_X \otimes \sigma_B) - \log \frac{3^3 \varepsilon}{\mu^3}, \end{aligned}$$

and Theorem 9, which yields

$$\begin{aligned} \max_{\sigma_B} D_h^{1 - \varepsilon^2 - \mu}(\rho_{XB} \| 1_X \otimes \sigma_B) \\ \geq D_h^{1 - \varepsilon^2 - \mu - \delta}(\rho_{XB} \| 1_X \otimes \rho_B) - \log \frac{2^3 \varepsilon}{\delta^2}. \end{aligned}$$

Combining the above relations, we obtain

$$\begin{aligned} -D_{\max}^{\varepsilon - \eta}(\rho_{XB} \| 1_X \otimes \rho_B) - \log \frac{2^3}{\eta^4} \\ \leq \ell^\varepsilon(X|B)_\rho \\ \leq \max_{\sigma_B} -D_{\max}^\varepsilon(\rho_{XB} \| 1_X \otimes \sigma_B) \\ \leq \max_{\sigma_B} -D_h^{1 - \varepsilon^2 - \mu}(\rho_{XB} \| 1_X \otimes \sigma_B) + \log \frac{3^3 \varepsilon}{\mu^3} \\ \leq -D_h^{1 - \varepsilon^2 - \mu - \delta}(\rho_{XB} \| 1_X \otimes \rho_B) + \log \frac{2^3 3^3 \varepsilon^2}{\delta^2 \mu^3}. \end{aligned} \quad (34)$$

We now consider the i.i.d. asymptotic setting with ρ_{XB}^n and its marginal ρ_B^n . Again, note that, $\lambda(\rho_B^n) = n\lambda(\rho_B)$ is linear concerning n , and, thus, $\log \theta(\rho_B^n)$ is of the order $O(\log n)$. Furthermore, the choice $\eta = \mu = \delta = 1/\sqrt{n}$ ensures that the additive terms are of the order $O(\log n)$.

This yields the following expansion due to (31) and (32).

Corollary 16: For any CQ state ρ_{XB} and any $0 < \varepsilon < 1$, we have the following asymptotic characterization for large n :

$$\ell^\varepsilon(X^n|B^n) = nH(X|B) + \sqrt{n} s(X|B) \Phi^{-1}(\varepsilon^2) + O(\log n).$$

We employed the conditional entropy

$$H(X|B)_\rho := -D(\rho_{XB} \| 1_X \otimes \rho_B)$$

as well as $s(A|B)_\rho := s(\rho_{AB} \| 1_A \otimes \rho_B)$.

VII. FINITE BLOCK LENGTH ANALYSIS

Our results of the previous section also directly imply bounds for finite block lengths, i.e. computable bounds on the operational quantities for fixed, large n . To get such bounds, we simply carefully combine the results presented above, which yields the following.

Theorem 17: Let ρ_{XB} be a CQ state and $0 < \varepsilon < 1$ be fixed. We use $s = s(X|B)_\rho$, $r = r(P_{\rho_{XB}, \rho_B} \| Q_{\rho_{XB}, \rho_B})$ and $\lambda = \lambda(\rho_B)$.⁶ Moreover, let $0 < \xi_0 < \min\{\varepsilon, 1 - \varepsilon\}$. Then, for any $n > \frac{C^2 r^6}{\xi_0^2 s^6}$, we have

$$\begin{aligned} \sup_{\xi_0 \leq \xi < 1 - \varepsilon} \left\{ -\sqrt{n} s \Phi^{-1}(\varepsilon + \xi) - \log \frac{2^9 \varepsilon \lceil n\lambda \rceil}{(\xi - \frac{C r^3}{\sqrt{n} s^3})^4 (1 - \varepsilon - \xi)} \right\} \\ \leq m^\varepsilon(X^n|B^n)_{\rho^n} - nH(X|B)_\rho \\ \leq \inf_{\xi_0 \leq \xi < \varepsilon} \left\{ -\sqrt{n} s \Phi^{-1}(\varepsilon - \xi) + \log \frac{2^2 3^3 \varepsilon \lceil n\lambda \rceil}{(\xi - \frac{C r^3}{\sqrt{n} s^3})^3} \right\}. \end{aligned}$$

⁶Note that, generally, $r(P_{\rho_{XB}, \rho_B} \| Q_{\rho_{XB}, \rho_B}) \neq r(\rho_{XB} \| 1_X \otimes \rho_B)$.

Furthermore, let $0 < \xi_1 < \min\{\varepsilon^2, 1 - \varepsilon^2\}$. Then, for any $n > \frac{C^2 r^6}{\xi_1^2 s^6}$, we have

$$\begin{aligned} & \sup_{\xi_1 \leq \xi < \varepsilon^2} \left\{ +\sqrt{n} s \Phi^{-1}(\varepsilon^2 - \xi) - \log \frac{5^5 \varepsilon^4 [n\lambda]}{(\xi - \frac{Cr^3}{\sqrt{ns^3}})^5 (1 - \varepsilon)} \right\} \\ & \leq \ell^\varepsilon(X^n | B^n)_{\rho^n} - n H(X | B)_\rho \\ & \leq \inf_{\xi_1 \leq \xi < 1 - \varepsilon^2} \left\{ +\sqrt{n} s \Phi^{-1}(\varepsilon^2 + \xi) + \log \frac{2^8 3^6 \varepsilon^2 [n\lambda]}{(\xi - \frac{Cr^3}{\sqrt{ns^3}})^6} \right\}. \end{aligned}$$

The remaining optimization over ξ is can be performed numerically. Note that any ξ in the required range gives valid lower and upper bounds on the operational quantities. Moreover, the asymptotic statement can be recovered when choosing ξ , ξ_0 and ξ_1 proportional to $1/\sqrt{n}$.

Proof: To get the first statement, we bound (33) using Theorem 14. This yields

$$\begin{aligned} & -D_s^{\varepsilon+\delta}(P_{\rho_{XB}, \rho_B} \| Q_{\rho_{XB}, \rho_B}) - \log \frac{2^8 \varepsilon \theta(\rho_B^n)}{\delta^4 (1 - \varepsilon - \delta)} \\ & \leq m^\varepsilon(X^n | B^n)_{\rho^n} \\ & \leq -D_s^{\varepsilon-\delta-\eta}(P_{\rho_{XB}, \rho_B} \| Q_{\rho_{XB}, \rho_B}) + \log \frac{2^3 \varepsilon \theta(\rho_B^n)}{\eta^2 \delta}. \end{aligned}$$

We further bound $\theta(\rho_B^n) \leq 2[n\lambda]$ and choose $\eta = 2\delta$. The Berry-Esseen Theorem (29) then gives us the expected bounds when we substitute ξ in the argument of Φ^{-1} . Note also that the parameter ξ can still be optimized over.

Similarly, bounding (34) using Theorem 14 yields

$$\begin{aligned} & -D_s^{1-(\varepsilon-\eta)^2+\delta}(P_{\rho_{XB}, \rho_B} \| Q_{\rho_{XB}, \rho_B}) - \log \frac{2^3 \theta(\rho_B^n)}{\eta^4 \delta (1 - \varepsilon)} \\ & \leq \ell^\varepsilon(X^n | B^n)_{\rho^n} \\ & \leq -D_s^{1-\varepsilon^2-\mu-\eta-\delta}(P_{\rho_{XB}, \rho_B} \| Q_{\rho_{XB}, \rho_B}) + \log \frac{2^3 3^3 \varepsilon^2 \theta(\rho_B^n)}{\mu^3 \eta^2 \delta}. \end{aligned} \quad (35)$$

The expression can be simplified using $\theta(\rho_B^n) \leq 2[n\lambda]$. Then, the upper bound follows by choosing $\mu = 3\delta$ and $\eta = 2\delta$ and substituting ξ as above after applying (29).

The optimization leading to the lower bound is a bit more involved. However, it is easy to verify that the choices $\eta = \frac{2\zeta}{5\varepsilon}$ and $\delta = \frac{\zeta}{5} + \frac{4\zeta^2}{25\varepsilon^2}$ lead to $(\varepsilon - \eta)^2 - \delta = \varepsilon^2 - \zeta$. Then, further bounding $\delta \geq \frac{\eta\varepsilon}{2}$, and substituting ξ as above, we arrive at the desired bound. ■

VIII. RELATION TO QUANTUM INFORMATION SPECTRUM

In the framework of the quantum information spectrum method, one treats general sequences of quantities $\vec{\alpha} = \{\alpha_n\}_{n=1}^\infty$ and investigates their asymptotic behavior. Given a sequence of Hilbert spaces, \mathcal{H} , and two sequences of states, $\vec{\rho}$ and $\vec{\sigma}$, such that $\rho_n, \sigma_n \in \mathcal{S}(\mathcal{H}_n)$ for all n , the quantum information spectrum is defined as [12]

$$\begin{aligned} \underline{D}(\varepsilon | \vec{\rho} | \vec{\sigma}) &:= \sup \{ R \in \mathbb{R} \mid \limsup_{n \rightarrow \infty} \text{tr } \rho_n \{ \rho_n \leq 2^{nR} \sigma_n \} \leq \varepsilon \}, \\ \overline{D}(\varepsilon | \vec{\rho} | \vec{\sigma}) &:= \inf \{ R \in \mathbb{R} \mid \liminf_{n \rightarrow \infty} \text{tr } \rho_n \{ \rho_n \leq 2^{nR} \sigma_n \} \geq \varepsilon \} \\ &= \sup \{ R \in \mathbb{R} \mid \liminf_{n \rightarrow \infty} \text{tr } \rho_n \{ \rho_n \leq 2^{nR} \sigma_n \} < \varepsilon \}. \end{aligned}$$

Our goal is to show that this can be expressed in terms of the entropic quantity D_s^ε that was used in the previous sections. For this purpose, we need the following lemma.

Lemma 18: Let \vec{g} be a sequence of monotonically increasing functions and define $f_n(\varepsilon) := \sup \{ R \mid g_n(R) \leq \varepsilon \}$. Then,

$$\begin{aligned} & \sup_{\vec{\varepsilon}} \left\{ \liminf_{n \rightarrow \infty} f_n(\varepsilon_n) \mid \limsup_{n \rightarrow \infty} \varepsilon_n \leq \varepsilon \right\} \\ & = \sup \{ R \in \mathbb{R} \mid \limsup_{n \rightarrow \infty} g_n(R) \leq \varepsilon \}, \quad \text{and} \end{aligned} \quad (36)$$

$$\begin{aligned} & \sup_{\vec{\varepsilon}} \left\{ \liminf_{n \rightarrow \infty} f_n(\varepsilon_n) \mid \liminf_{n \rightarrow \infty} \varepsilon_n < \varepsilon \right\} \\ & = \sup \{ R \in \mathbb{R} \mid \liminf_{n \rightarrow \infty} g_n(R) < \varepsilon \}. \end{aligned} \quad (37)$$

Proof: We prove Eq. (36) and simply note that Eq. (37) can be shown analogously.

By definition of the supremum, for any $\delta > 0$, there exists a real R' satisfying $\limsup_{n \rightarrow \infty} g_n(R') \leq \varepsilon$ and

$$\sup \{ R \in \mathbb{R} \mid \limsup_{n \rightarrow \infty} g_n(R) \leq \varepsilon \} < R' + \delta.$$

We now define a sequence $\vec{\varepsilon} = \{\varepsilon_n\}_{n=1}^\infty$ using $\varepsilon_n = g_n(R')$. Then, we have $\limsup_{n \rightarrow \infty} \varepsilon_n \leq \varepsilon$ and, since $R' \leq f_n(g_n(R')) = f_n(\varepsilon_n)$ for all n by definition of f_n , we find $R' \leq \liminf_{n \rightarrow \infty} f_n(\varepsilon_n)$. Hence,

$$R' \leq \sup_{\vec{\varepsilon}} \left\{ \liminf_{n \rightarrow \infty} f_n(\varepsilon_n) \mid \limsup_{n \rightarrow \infty} \varepsilon_n \leq \varepsilon \right\}.$$

Conversely, there exists a sequence $\vec{\varepsilon}'$ satisfying $\limsup_{n \rightarrow \infty} \varepsilon'_n \leq \varepsilon$ and

$$\sup_{\vec{\varepsilon}'} \left\{ \liminf_{n \rightarrow \infty} f_n(\varepsilon_n) \mid \limsup_{n \rightarrow \infty} \varepsilon_n \leq \varepsilon \right\} < \liminf_{n \rightarrow \infty} f_n(\varepsilon'_n) + \delta.$$

We now define $R = \liminf_{n \rightarrow \infty} f_n(\varepsilon'_n)$. Then, by definition of the limit inferior, there exists an integer n_0 such that $R - \delta < f_n(\varepsilon'_n)$ for $n \geq n_0$. Hence, $g_n(R - \delta) \leq \varepsilon'_n$ for $n \geq n_0$ and, thus, $\limsup_{n \rightarrow \infty} g_n(R - \delta) \leq \limsup_{n \rightarrow \infty} \varepsilon'_n \leq \varepsilon$. Thus,

$$\begin{aligned} \liminf_{n \rightarrow \infty} f_n(\varepsilon'_n) &\leq \sup \{ R \mid \limsup_{n \rightarrow \infty} g_n(R - \delta) \leq \varepsilon \} \\ &= \sup \{ R \mid \limsup_{n \rightarrow \infty} g_n(R) \leq \varepsilon \} + \delta. \end{aligned}$$

Since we may choose δ arbitrarily small, the above inequalities establish equality in (36). ■

We now employ Lemma 18 using the sequence of functions $g_n(R) = \text{tr } \rho_n \{ \rho_n \leq e^{nR} \sigma_n \}$, and, hence, $f_n(\varepsilon) = D_s^\varepsilon(\rho_n \| \sigma_n)$ by definition. This yields the following equalities.

$$\begin{aligned} \underline{D}(\varepsilon | \vec{\rho} | \vec{\sigma}) &= \sup_{\vec{\varepsilon}} \left\{ \liminf_{n \rightarrow \infty} \frac{1}{n} D_s^{\varepsilon_n}(\rho_n \| \sigma_n) \mid \limsup_{n \rightarrow \infty} \varepsilon_n \leq \varepsilon \right\}, \\ \overline{D}(\varepsilon | \vec{\rho} | \vec{\sigma}) &= \sup_{\vec{\varepsilon}} \left\{ \liminf_{n \rightarrow \infty} \frac{1}{n} D_s^{\varepsilon_n}(\rho_n \| \sigma_n) \mid \liminf_{n \rightarrow \infty} \varepsilon_n < \varepsilon \right\}. \end{aligned}$$

This shows that relative entropy D_s constitutes an entropic version of the information spectrum \overline{D} and \underline{D} .

Together with the hierarchy derived in the previous section, this allows us to relate various information quantities to the quantum information spectrum. As an example, the following operational quantities are used to analyze quantum hypothesis

testing [12]. The asymptotic achievability is given by

$$\begin{aligned} B(\varepsilon|\vec{\rho}||\vec{\sigma}) &:= \sup_{\vec{Q}} \left\{ \liminf_{n \rightarrow \infty} \frac{-1}{n} \log \langle \sigma_n, Q_n \rangle \mid \bar{\varepsilon}(\vec{Q}) \leq \varepsilon \right\} = \\ &\sup \left\{ R \in \mathbb{R} \mid \exists \vec{Q} : \liminf_{n \rightarrow \infty} \frac{-1}{n} \log \langle \sigma_n, Q_n \rangle \geq R \wedge \bar{\varepsilon}(\vec{Q}) \leq \varepsilon \right\} \\ &= \sup_{\vec{\varepsilon}} \left\{ \liminf_{n \rightarrow \infty} \frac{1}{n} D_h^{\varepsilon_n}(\rho_n \| \sigma_n) \mid \limsup_{n \rightarrow \infty} \varepsilon_n \leq \varepsilon \right\}, \end{aligned}$$

where we used $\bar{\varepsilon}(\vec{Q}) = \limsup_{n \rightarrow \infty} \langle \rho_n, 1 - Q_n \rangle$. On the other hand, the asymptotic converse is described by

$$\begin{aligned} B^\dagger(\varepsilon|\vec{\rho}||\vec{\sigma}) &:= \sup_{\vec{Q}} \left\{ \liminf_{n \rightarrow \infty} \frac{-1}{n} \log \langle \sigma_n, Q_n \rangle \mid \underline{\varepsilon}(\vec{Q}) < \varepsilon \right\} = \\ &\inf \left\{ R \in \mathbb{R} \mid \forall \vec{Q} : \liminf_{n \rightarrow \infty} \frac{-1}{n} \log \langle \sigma_n, Q_n \rangle \geq R \implies \underline{\varepsilon}(\vec{Q}) \geq \varepsilon \right\} \\ &= \sup_{\vec{\varepsilon}} \left\{ \liminf_{n \rightarrow \infty} \frac{1}{n} D_h^{\varepsilon_n}(\rho_n \| \sigma_n) \mid \liminf_{n \rightarrow \infty} \varepsilon_n < \varepsilon \right\}, \end{aligned}$$

where we used $\underline{\varepsilon}(\vec{Q}) = \liminf_{n \rightarrow \infty} \langle \rho_n, 1 - Q_n \rangle$. The equalities with the expressions involving the one-shot entropy can be verified by choosing $\varepsilon_n = \langle \rho_n, 1 - Q_n \rangle$ for any sequence \vec{Q} . Conversely, for any sequence $\vec{\varepsilon}$ satisfying the constraint, we choose Q_n as the primal optimal solution for $D_h^{\varepsilon_n}(\rho \| \sigma)$.

Using Lemma 12, we can confirm the following result [12]:

$$B(\varepsilon|\vec{\rho}||\vec{\sigma}) = \underline{D}(\varepsilon|\vec{\rho}||\vec{\sigma}) \quad \text{and} \quad B^\dagger(\varepsilon|\vec{\rho}||\vec{\sigma}) = \overline{D}(\varepsilon|\vec{\rho}||\vec{\sigma}).$$

Furthermore, Theorem 14 and Lemma 12 together imply that as long as the number of distinct eigenvalues in σ_n or the logarithm of the minimum eigenvalue in σ_n grows at most polynomially in n , we get the following, novel, relations:

$$B(\varepsilon|\vec{\rho}||\vec{\sigma}) = B(\varepsilon|\vec{P}_0||\vec{P}_1) \quad \text{and} \quad B^\dagger(\varepsilon|\vec{\rho}||\vec{\sigma}) = B^\dagger(\varepsilon|\vec{P}_0||\vec{P}_1),$$

where \vec{P}_i is the sequence of classical distributions $\{P_{i,\rho_n,\sigma_n}\}_n$ as defined in Section VI. The latter quantities can be bounded further using results from classical hypothesis testing.

Furthermore, we want to point out that our analysis can be used to extend results by Datta and Renner [9] relating the information spectrum for $\varepsilon \in \{0, 1\}$ and smooth min- and max-entropies to arbitrary $0 < \varepsilon < 1$. If the eigenvalues of σ_n satisfy the condition of the previous paragraph, then (18) and (19) imply the following results:

$$\begin{aligned} \underline{D}(\varepsilon|\vec{\rho}||\vec{\sigma}) &= \sup_{\vec{\varepsilon}} \left\{ \liminf_{n \rightarrow \infty} \frac{1}{n} D_{\max}^{\sqrt{1-\varepsilon_n}}(\rho_n \| \sigma_n) \mid \limsup_{n \rightarrow \infty} \varepsilon_n \leq \varepsilon \right\}, \\ \overline{D}(\varepsilon|\vec{\rho}||\vec{\sigma}) &= \sup_{\vec{\varepsilon}} \left\{ \liminf_{n \rightarrow \infty} \frac{1}{n} D_{\max}^{\sqrt{1-\varepsilon_n}}(\rho_n \| \sigma_n) \mid \liminf_{n \rightarrow \infty} \varepsilon_n < \varepsilon \right\}, \end{aligned}$$

which can be readily specialized to conditional entropies.

Finally, we want to point out that the sequences of rates, $\{\frac{1}{n} m^{\varepsilon_n}(X|B)_{\rho_n}\}_n$ and $\{\frac{1}{n} \ell^{\varepsilon_n}(X|B)_{\rho_n}\}_n$, can be expressed asymptotically using the information spectrum method analogously to the case of hypothesis testing. Our results then show that their asymptotics are equal to the asymptotics of $\{\frac{1}{n} H_h^{\varepsilon_n}(X|B)_{\rho_n}\}_n$ and $\{\frac{1}{n} H_{\min}^{\varepsilon_n}(X|B)_{\rho_n}\}_n$, respectively. Furthermore, if the abovementioned conditions on the eigenvalues are satisfied, these expressions correspond to the information spectrum, \underline{D} and \overline{D} . This is discussed in detail in Appendix B.

IX. CONCLUSION AND DISCUSSION

We characterize both source compression and randomness extraction with quantum side information using one-shot entropies in such a way that the second order asymptotics of these tasks can be recovered. This result improves on previous characterizations of these quantities that were only shown to converge in the first order.

We want to point out the relation of our result to the smooth entropy framework that has recently been employed to characterize various quantum information theoretic tasks in the one-shot setting. Such characterizations allow to recover the correct asymptotic behavior in the first order, and that is often taken as a sufficient reason to call them “tight”. However, we stress that a first order analysis is independent of the required security or error parameter, ε .⁷ Hence, a characterization with H_{\min}^{ε} is equivalent to a characterization with $H_{\min}^{2\varepsilon}$, $H_{\min}^{\sqrt{\varepsilon}}$ or even $H_{\min}^{1-\varepsilon}$ in the first order—a freedom that is often used extensively to prove these results. In the second order, however, the above quantities behave very differently. Hence, tightness in the second order requires a more precise analysis of the one-shot problem, resulting in a characterization of the operational quantities in terms of $H_{\min}^{\varepsilon \pm \eta}$ plus terms that grow at most proportional to $\log \frac{1}{\eta}$ when $\eta \rightarrow 0$.

It appears that such a characterization is only possible in terms of a carefully chosen one-shot entropy, depending on the task at hand. We show that the hypothesis testing entropy, $H_h^{\varepsilon \pm \eta}$, allows a tight one-shot characterization of source compression, while the smooth min-entropy, $H_{\min}^{\varepsilon \pm \eta}$ takes the respective role for randomness extraction when the secrecy criterion is given in terms of the purified distance. In conclusion, we do not expect that a single one-shot entropy is sufficient to characterize all relevant tasks such that the correct second order asymptotics can be recovered.

Finally, we established in Section VIII that the behavior of the asymptotic information spectrum of a task—both its direct and converse part—can be expressed as an appropriate limit of the respective one-shot quantity. Hence, a thorough analysis of the one-shot quantity leads also to the understanding of the behavior of the asymptotic information spectrum.

ACKNOWLEDGMENTS

MT would like to thank Frédéric Dupuis and Joseph M. Renes for stimulating discussions about the hypothesis testing entropy and the importance of finite block length analysis. This work is supported by the National Research Foundation and the Ministry of Education of Singapore. MH is partially supported by a MEXT Grant-in-Aid for Young Scientists (A) No. 20686026, Grant-in-Aid for Scientific Research (A) No. 23246071, and by the National Institute of Information and Communication Technology (NICT), Japan.

APPENDIX A

EXAMPLE OF FINITE BLOCK LENGTH ANALYSIS: EAVESDROPPING ON PAULI CHANNEL

We consider the state that results when transmitting either $|0\rangle$ or $|1\rangle$ through the complementary channel to a Pauli

⁷In the contrary, such an analysis is expected to yield the same first order asymptotic expansion for all $0 < \varepsilon < 1$.

channel with a phase error $p < \frac{1}{2}$ that is independent of the bit flip error. The resulting state is

$$\rho_{XB} = \frac{1}{2} \sum_{x=0}^1 |x\rangle\langle x| \otimes |\phi^x\rangle\langle\phi^x|, \quad \text{where} \\ |\phi^x\rangle = \sqrt{p}|0\rangle + (-1)^x \sqrt{1-p}|1\rangle.$$

Moreover, we note that the non-trivial part of the Nussbaum-Szkoła distribution for this state reads

$$P = P_{\rho_{XB}, 1_X \otimes \rho_B} = \left\{ \frac{p}{2}, \frac{p}{2}, \frac{1-p}{2}, \frac{1-p}{2} \right\} \quad \text{and} \\ Q = Q_{\rho_{XB}, 1_X \otimes \rho_B} = \{p^2, p^2, (1-p)^2, (1-p)^2\}.$$

We are interested in how much randomness can be extracted from n i.i.d. copies of this source for finite n , i.e. we want to find bounds on $\ell^\varepsilon(X^n|B^n)$. We first bound this in terms of the classical information spectrum. For any choices of $\xi_1, \xi_2 \in (0, 1)$, we write using (35),

$$\begin{aligned} -D_s^{1-\varepsilon^2(1-\xi_1)}(P^n\|Q^n) - \log \frac{5^5 \lceil n\lambda \rceil}{\xi_1^5 \varepsilon^6 (1-\varepsilon)} \\ \leq \ell^\varepsilon(X^n|B^n) \\ \leq -D_s^{1-\varepsilon^2(1+\xi_2)}(P^n\|Q^n) + \log \frac{2^8 3^6 \lceil n\lambda \rceil}{\xi_2^6 \varepsilon^{10}} \end{aligned} \quad (38)$$

Then, we note that $D_s^\varepsilon(P^{\otimes n}\|Q^{\otimes n})$ can be evaluated precisely as follows. First, we recall (28) and write

$$D_s^\varepsilon(P^n\|Q^n) = \sup \left\{ R \mid P^n \left\{ \sum_i Z_i \leq R \right\} \leq \varepsilon \right\},$$

where $Z_i = \log P - \log Q$ is a random variable that takes value $\log \frac{1}{2p}$ with probability p and value $\log \frac{1}{2(1-p)}$ with probability $1-p$. We rescale this into a Bernoulli trial $B_i = (Z_i - \log \frac{1}{2(1-p)}) (\log \frac{1-p}{p})^{-1}$ and find

$$\begin{aligned} P^n \left\{ \sum_i Z_i \leq R \right\} \\ = P^n \left\{ \sum_i B_i \leq \underbrace{\left(R - n \log \frac{1}{2(1-p)} \right) \left(\log \frac{1-p}{p} \right)^{-1}}_{\tilde{R}} \right\}, \end{aligned}$$

where we used that $\log \frac{1-p}{p}$ is positive for $p < \frac{1}{2}$. Hence,

$$\begin{aligned} D_s^\varepsilon(P^n\|Q^n) \\ = \sup \left\{ \tilde{R} \mid P^n \left\{ \sum_i B_i \leq \tilde{R} \right\} \leq \varepsilon \right\} \log \frac{1-p}{p} + n \log \frac{1}{2(1-p)} \\ = \max \left\{ k \in \mathbb{N} \mid F(k-1; n, p) \leq \varepsilon \right\} \log \frac{1-p}{p} + n \log \frac{1}{2(1-p)}, \end{aligned}$$

where $F(\cdot; n, p)$ is the cumulative distribution function for the binomial distribution and the remaining optimization can be evaluated numerically. Combining this with (38), we can thus evaluate direct and converse bounds on the extractable randomness in Fig. 1.

APPENDIX B THE INFORMATION SPECTRUM OF SOURCE COMPRESSION AND RANDOMNESS EXTRACTION

Here, we treat source compression and randomness extraction using the information spectrum method. That is, we focus on the asymptotic operational quantities for general sequence of information sources $\vec{\rho}_{XB} := \{\rho_{XB,n}\}_n$.

We define

$$\begin{aligned} m(\varepsilon|X|B|\vec{\rho}) \\ := \inf_{\vec{P}} \{ \limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathcal{M}(\mathcal{P}_n) \mid \limsup_{n \rightarrow \infty} p_{\text{err}}(\mathcal{P}_n, \rho_{XB,n}) \leq \varepsilon \} \\ = \inf_{\vec{\varepsilon}} \{ \limsup_{n \rightarrow \infty} \frac{1}{n} m^{\varepsilon_n}(X|B)_{\rho_n} \mid \limsup_{n \rightarrow \infty} \varepsilon_n \leq \varepsilon \} \\ m^\dagger(\varepsilon|X|B|\vec{\rho}) \\ := \inf_{\vec{P}} \{ \limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathcal{M}(\mathcal{P}_n) \mid \liminf_{n \rightarrow \infty} p_{\text{err}}(\mathcal{P}_n, \rho_{XB,n}) < \varepsilon \} \\ = \inf_{\vec{\varepsilon}} \{ \limsup_{n \rightarrow \infty} \frac{1}{n} m^{\varepsilon_n}(X|B)_{\rho_n} \mid \liminf_{n \rightarrow \infty} \varepsilon_n < \varepsilon \} \\ \ell(\varepsilon|X|B|\vec{\rho}) \\ := \inf_{\vec{P}} \{ \liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathcal{Z}(\mathcal{P}_n) \mid \limsup_{n \rightarrow \infty} d_{\text{sec}}(\mathcal{P}_n, \rho_{XB,n}) \leq \varepsilon \} \\ = \inf_{\vec{\varepsilon}} \{ \liminf_{n \rightarrow \infty} \frac{1}{n} \ell^{\varepsilon_n}(X|B)_{\rho_n} \mid \limsup_{n \rightarrow \infty} \varepsilon_n \leq \varepsilon \} \\ \ell^\dagger(\varepsilon|X|B|\vec{\rho}) \\ := \inf_{\vec{P}} \{ \liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathcal{Z}(\mathcal{P}_n) \mid \liminf_{n \rightarrow \infty} d_{\text{sec}}(\mathcal{P}_n, \rho_{XB,n}) < \varepsilon \} \\ = \inf_{\vec{\varepsilon}} \{ \liminf_{n \rightarrow \infty} \frac{1}{n} \ell^{\varepsilon_n}(X|B)_{\rho_n} \mid \liminf_{n \rightarrow \infty} \varepsilon_n \leq \varepsilon \} \end{aligned}$$

In order to characterize these quantities, we define the asymptotic quantum conditional entropies:

$$\begin{aligned} \overline{H}(\varepsilon|X|B|\vec{\rho}) \\ := \sup \{ R \in \mathbb{R} \mid \liminf_{n \rightarrow \infty} \text{tr} \rho_{XB,n} \{ \rho_{XB,n} > 2^{-nR} \rho_{B,n} \} < \varepsilon \} \\ \underline{H}(\varepsilon|X|B|\vec{\rho}) \\ := \sup \{ R \in \mathbb{R} \mid \liminf_{n \rightarrow \infty} \text{tr} \rho_{XB,n} \{ \rho_{XB,n} > 2^{-nR} \rho_{B,n} \} \leq \varepsilon \}. \end{aligned}$$

Using the classical distribution, we can define another kind of asymptotic quantum conditional entropies:

$$\begin{aligned} \overline{H}_c(\varepsilon|X|B|\vec{\rho}) \\ := \sup \{ R \in \mathbb{R} \mid \liminf_{n \rightarrow \infty} P_n \{ P_n > 2^{-nR} Q_n \} < \varepsilon \} \\ \underline{H}_c(\varepsilon|X|B|\vec{\rho}) \\ := \sup \{ R \in \mathbb{R} \mid \liminf_{n \rightarrow \infty} P_n \{ P_n > 2^{-nR} Q_n \} \leq \varepsilon \}, \end{aligned}$$

where $P_n := P_{\rho_{XB,n}, \rho_{B,n}}$ and $Q_n := Q_{\rho_{XB,n}, \rho_{B,n}}$.

Similarly, we can define the asymptotic smooth conditional min-entropy

$$\begin{aligned} \overline{H}_{\min}(\varepsilon|X|B|\vec{\rho}) &:= \sup_{\vec{\varepsilon}} \{ \liminf_{n \rightarrow \infty} \frac{1}{n} H_{\min}^{\varepsilon_n}(X|B)_{\rho_n} \mid \liminf_{n \rightarrow \infty} \varepsilon_n < \varepsilon \} \\ \underline{H}_{\min}(\varepsilon|X|B|\vec{\rho}) &:= \sup_{\vec{\varepsilon}} \{ \liminf_{n \rightarrow \infty} \frac{1}{n} H_{\min}^{\varepsilon_n}(X|B)_{\rho_n} \mid \limsup_{n \rightarrow \infty} \varepsilon_n \leq \varepsilon \}. \end{aligned}$$

Applying Lemma 18 to the case when $f_n(\varepsilon) = -\frac{1}{n} D_s^{1-\varepsilon}(\rho_{XB,n} \| \rho_{B,n})$ and $g_n(R) = \text{tr} \rho_{XB,n} \{ \rho_{XB,n} >$

$2^{-nR}\rho_{B,n}\}$, we obtain

$$\begin{aligned} & \overline{H}(\varepsilon|X|B|\vec{\rho}) \\ &= \sup_{\varepsilon} \left\{ \liminf_{n \rightarrow \infty} -\frac{1}{n} D_s^{1-\varepsilon_n}(\rho_{XB,n} \|\rho_{B,n}) \mid \liminf_{n \rightarrow \infty} \varepsilon_n < \varepsilon \right\} \end{aligned} \quad (39)$$

$$\begin{aligned} & \underline{H}(\varepsilon|X|B|\vec{\rho}) \\ &= \sup_{\varepsilon} \left\{ \liminf_{n \rightarrow \infty} -\frac{1}{n} D_s^{1-\varepsilon_n}(\rho_{XB,n} \|\rho_{B,n}) \mid \limsup_{n \rightarrow \infty} \varepsilon_n \leq \varepsilon \right\}. \end{aligned} \quad (40)$$

The same type inequality holds for $\overline{H}_c(\varepsilon|X|B|\vec{\rho})$ and $\underline{H}_c(\varepsilon|X|B|\vec{\rho})$.

Now, we substitute ε_n and $\frac{1}{n}$ into ε and η in Theorem 9, and substitute ε_n and $\frac{1}{n}$ into ε and δ in Lemma 12. Due to relations (39) and (40),

$$m(\varepsilon|X|B|\vec{\rho}) = \overline{H}(1 - \varepsilon|X|B|\vec{\rho}) \quad (41)$$

holds for $0 < \varepsilon \leq 1$ and

$$m^\dagger(\varepsilon|X|B|\vec{\rho}) = \underline{H}(1 - \varepsilon|X|B|\vec{\rho}) \quad (42)$$

holds for $0 \leq \varepsilon < 1$.

Similarly, we substitute ε_n and $\frac{1}{n}$ into ε and η in Theorem 8. Then,

$$\ell(\varepsilon|X|B|\vec{\rho}) = \overline{H}_{\min}(\varepsilon|X|B|\vec{\rho}) \quad (43)$$

holds for $0 < \varepsilon \leq 1$ and

$$\ell^\dagger(\varepsilon|X|B|\vec{\rho}) = \underline{H}_{\min}(\varepsilon|X|B|\vec{\rho}) \quad (44)$$

holds for $0 \leq \varepsilon < 1$.

In the following, we assume the above mentioned conditions on the eigenvalues. That is, the number of distinct eigenvalues in σ_n or the logarithm of the minimum eigenvalue in σ_n is assumed to grow at most polynomially in n . Then, combining relations (39) and (40), Lemma 12, and Theorem 14, we obtain

$$\overline{H}(\varepsilon^2|X|B|\vec{\rho}) = \overline{H}_c(\varepsilon^2|X|B|\vec{\rho}) = \overline{H}_{\min}(\varepsilon|X|B|\vec{\rho}) \quad (45)$$

$$\underline{H}(\varepsilon^2|X|B|\vec{\rho}) = \underline{H}_c(\varepsilon^2|X|B|\vec{\rho}) = \underline{H}_{\min}(\varepsilon|X|B|\vec{\rho}). \quad (46)$$

Note that these equations hold without the above mentioned conditions on the eigenvalues in the commutative case.

REFERENCES

- [1] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Channel Coding Rate in the Finite Blocklength Regime," *IEEE Trans. on Inf. Theory*, vol. 56, no. 5, pp. 2307–2359, May 2010.
- [2] M. Hayashi, "Information Spectrum Approach to Second-Order Coding Rate in Channel Coding," *IEEE Trans. on Inf. Theory*, vol. 55, no. 11, pp. 4947–4966, Nov. 2009.
- [3] I. Kontoyiannis, "Second-Order Noiseless Source Coding Theorems," *IEEE Trans. on Inf. Theory*, vol. 43, no. 4, pp. 1339–1341, Jul. 1997.
- [4] M. Hayashi, "Second-Order Asymptotics in Fixed-Length Source Coding and Intrinsic Randomness," *IEEE Trans. on Inf. Theory*, vol. 54, no. 10, pp. 4619–4637, Oct. 2008.
- [5] D. Baron, M. A. Khojastepour, and R. G. Baraniuk, "Redundancy Rates of Slepian-Wolf Coding," in *Proc. 42nd Annual Allerton Conf. on Comm., Control, and Computing*, 2004.
- [6] V. Y. F. Tan and O. Kosut, "The Dispersion of Slepian-Wolf Coding," in *Proc. IEEE ISIT*, 2012.
- [7] M. Hayashi, *Quantum Information — An Introduction*. Springer, 2006.
- [8] Y. Polyanskiy, "Channel Coding: Non-Asymptotic Fundamental Limits," Ph.D. dissertation, Princeton University, Nov. 2010.
- [9] N. Datta and R. Renner, "Smooth Entropies and the Quantum Information Spectrum," *IEEE Trans. on Inf. Theory*, vol. 55, no. 6, pp. 2807–2815, Jun. 2009.
- [10] R. Renner, "Security of Quantum Key Distribution," Ph.D. dissertation, ETH Zurich, Dec. 2005. [Online]. Available: <http://arxiv.org/abs/quant-ph/0512258>
- [11] M. Hayashi and H. Nagaoka, "General Formulas for Capacity of Classical-Quantum Channels," *IEEE Trans. on Inf. Theory*, vol. 49, no. 7, pp. 1753–1768, Jul. 2003.
- [12] H. Nagaoka and M. Hayashi, "An Information-Spectrum Approach to Classical and Quantum Hypothesis Testing for Simple Hypotheses," *IEEE Trans. on Inf. Theory*, vol. 53, no. 2, pp. 534–549, Feb. 2007.
- [13] T. S. Han and S. Verdú, "Approximation Theory of Output Statistics," in *Proc. IEEE ISIT*. IEEE, 1993, pp. 153–153.
- [14] T. S. Han, *Information-Spectrum Methods in Information Theory*. Springer, 2002.
- [15] A. Rényi, "On Measures of Information and Entropy," in *Proc. Symp. on Math., Stat. and Probability*. Berkeley: University of California Press, 1961, pp. 547–561.
- [16] R. Renner and R. König, "Universally Composable Privacy Amplification Against Quantum Adversaries," in *Proc. TCC*, ser. LNCS, vol. 3378, Cambridge, USA, 2005, pp. 407–425.
- [17] M. Tomamichel, R. Colbeck, and R. Renner, "A Fully Quantum Asymptotic Equipartition Property," *IEEE Trans. on Inf. Theory*, vol. 55, no. 12, pp. 5840–5847, Dec. 2009.
- [18] —, "Duality Between Smooth Min- and Max-Entropies," *IEEE Trans. on Inf. Theory*, vol. 56, no. 9, pp. 4674–4681, Sep. 2010.
- [19] M. Tomamichel, "A Framework for Non-Asymptotic Quantum Information Theory," Ph.D. dissertation, ETH Zurich, Mar. 2012. [Online]. Available: <http://arxiv.org/abs/1203.2142>
- [20] J. M. Renes and R. Renner, "One-Shot Classical Data Compression With Quantum Side Information and the Distillation of Common Randomness or Secret Keys," *IEEE Trans. on Inf. Theory*, vol. 58, no. 3, pp. 1985–1991, Mar. 2012.
- [21] M. Berta, "Single-Shot Quantum State Merging," Master's Thesis, ETH Zurich, 2008.
- [22] F. Buscemi and N. Datta, "The Quantum Capacity of Channels With Arbitrarily Correlated Noise," *IEEE Trans. on Inf. Theory*, vol. 56, no. 3, pp. 1447–1460, Mar. 2010.
- [23] I. Devetak and A. Winter, "Classical Data Compression with Quantum Side Information," *Phys. Rev. A*, vol. 68, no. 4, Oct. 2003.
- [24] A. S. Holevo, "An Analog of the Theory of Statistical Decisions in Noncommutative Theory of Probability," *Trans. Moscow Math. Soc.*, vol. 26, pp. 133–149, 1972.
- [25] C. W. Helstrom, *Quantum Detection and Estimation Theory*. New York: Academic Press, 1976.
- [26] F. Hiai and D. Petz, "The Proper Formula for Relative Entropy and its Asymptotics in Quantum Probability," *Comm. Math. Phys.*, vol. 143, no. 1, pp. 99–114, Dec. 1991.
- [27] H. Nagaoka and T. Ogawa, "Strong converse and Stein's lemma in quantum hypothesis testing," *IEEE Trans. on Inf. Theory*, vol. 46, no. 7, pp. 2428–2433, Nov. 2000.
- [28] M. Hayashi, "Optimal Sequence of Quantum Measurements in the Sense of Stein's Lemma in Quantum Hypothesis Testing," *J. Phys. A: Math. Gen.*, vol. 35, no. 50, pp. 10 759–10 773, 2002.
- [29] M. Nussbaum and A. Szkola, "The Chernoff Lower Bound for Symmetric Quantum Hypothesis Testing," *Ann. Stat.*, vol. 37, no. 2, pp. 1040–1057, Apr. 2009.
- [30] L. Wang and R. Renner, "One-Shot Classical-Quantum Capacity and Hypothesis Testing," *Phys. Rev. Lett.*, vol. 108, no. 20, May 2012.
- [31] R. König, R. Renner, and C. Schaffner, "The Operational Meaning of Min- and Max-Entropy," *IEEE Trans. on Inf. Theory*, vol. 55, no. 9, pp. 4337–4347, Sep. 2009.
- [32] K. M. R. Audenaert, M. Mosonyi, and F. Verstraete, "Quantum state discrimination bounds for finite sample size," Apr. 2012.
- [33] K. Li, "Second Order Asymptotics for Quantum Hypothesis Testing," Aug. 2012.
- [34] N. Datta, "Min- and Max- Relative Entropies and a New Entanglement Monotone," *IEEE Trans. on Inf. Theory*, vol. 55, no. 6, pp. 2816–2826, 2009.
- [35] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.
- [36] J. L. Carter and M. N. Wegman, "Universal Classes of Hash Functions," *J. Comp. Syst. Sci.*, vol. 18, no. 2, pp. 143–154, Apr. 1979.
- [37] M. Ohya and D. Petz, *Quantum Entropy and Its Use*. Springer, 1993.

- [38] I. Tyurin, “An Improvement of Upper Estimates of the Constants in the Lyapunov Theorem,” *Russian Math. Surveys*, vol. 65, no. 3, pp. 201–202, 2010.